ARGUMENTS *for a* BETTER WORLD

Essays in Honor of Amartya Sen

VOLUME II

Society, Institutions, and Development

EDITED BY

Kaushik Basu and Ravi Kanbur

ARGUMENTS FOR A BETTER WORLD

ESSAYS IN HONOR OF AMARTYA SEN

This page intentionally left blank

ARGUMENTS FOR A BETTER WORLD

ESSAYS IN HONOR OF AMARTYA SEN

Volume II: Society, Institutions, and Development

Edited by

KAUSHIK BASU

and

RAVI KANBUR

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States
by Oxford University Press Inc., New York

First published 2009

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Arguments for a better world : essays in honor of Amartya Sen /
edited by Kaushik Basu Ravi Kanbur.
p. cm.
ISBN 978–0–19–923911–5 (hbk. : v. 1) – ISBN 978–0–19–923997–9 (hbk. : v. 2)
1. Welfare economics. 2. Institutional economics. 3. Economics–Moral and
ethical aspects. I. Basu, Kaushik. II. Kanbur, S. M. Ravi III. Sen, Amartya Kumar.
IV. Title:
Essays in honor of Amartya Sen.
HB846.094 2008
330–dc22 2008036360

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Italy by
Rotolito Lombarda S.p.A.

ISBN 978–0–19–923911–5 (Volume 1)
ISBN 978–0–19–923997–9 (Volume 2)
ISBN 978–0–19–923999–3 (Set)

1 3 5 7 9 10 8 6 4 2

Contents

PART IV SOCIETY, POLITICS AND HISTORY

List of Figures

List of Tables

List of Contributors

Bina Agarwal, University of Delhi

Isher Ahluwalia, Indian Council for Research on International Economic Relations

Montek S. Ahluwalia, Planning Commission, Government of India

Ingela Alger, Carleton University

Muhammad Asali, Columbia University

Amiya Kumar Bagchi, Institute of Development Studies, Kolkata

Pranab Bardhan, University of California, Berkeley

Kaushik Basu, Cornell University

Lourdes Benería, Cornell University

Sugata Bose, Harvard University

Lincoln C. Chen, China Medical Board

Martha Alter Chen, Harvard University

Kanchan Chopra, University of Delhi

Simon Dietz, London School of Economics and Political Science

Sakiko Fukuda-Parr, The New School

Jonathan Glover, King's College London

Cameron Hepburn, University of Oxford

Jane Humphries, University of Oxford

Rizwanul Islam, International Labour Office, Geneva

Ayesha Jalal, Tufts University

Mary Kaldor, London School of Economics and Political Science

Ravi Kanbur, Cornell University

Sunil Khilnani, Johns Hopkins University

Stephan Klasen, University of Göttingen

Jocelyn Kynch, retired from University of Swansea

Enrica Chiappero Martinetti, University of Pavia

Kirsty McNay, independent researcher

Martha C. Nussbaum, University of Chicago

Elinor Ostrom, Indiana University, Bloomington

Gustav Ranis, Yale University

Sanjay G. Reddy, Barnard College and Columbia University

Emma Samman, University of Oxford

Rehman Sobhan, Centre for Policy Dialogue, Bangladesh

Robert M. Solow, Massachusetts Institute of Technology

Nicholas Stern, London School of Economics and Political Science

Frances Stewart, University of Oxford

Ashutosh Varshney, University of Michigan

Sujata Visaria, Boston University

Jörgen W. Weibull, Stockholm School of Economics

This page intentionally left blank

INTRODUCTION

KAUSHIK BASU
RAVI KANBUR

When Amartya Sen received the Nobel Prize in Economics in 1998, there was much confusion in popular reportage about which of his contributions were being honored by the prize. Some wrote that he got the prize for his large contributions to development economics, some claimed this was recognition for his pioneering work on the history and causes of famines, and yet others suggested that he got it for the moral causes he espoused.

Sen got the prize for welfare economics. Part of the blame for the confusion in people's minds has to be directed at him, for it is caused by the truly remarkable range of his writings. He has indeed contributed to all the fields people thought he had got the prize for and he has been a relentless campaigner for moral causes, such as human rights, basic freedoms and capabilities and the guarantee of minimum living standards. What is even more remarkable is that he has continued to broaden the areas of his attention and inquiry since 1998. He has worked on Indian history and the politics of identity and violence, and has recently chaired the Commonwealth Commission on "Civil Paths to Peace". He can lay claim to being a social scientist in a way that few others can.

On 3 November 2008, Amartya Sen will turn 75; and we, as his one-time students and, subsequently, colleagues in the broader sense of the term, felt that it was a fitting occasion to celebrate his remarkable achievements. We wanted to do this by presenting him with a book that tries to push the frontiers further on the full range of academic themes and activist causes with which Sen has been associated. One thing was immediately evident—this would not be a small project. We decided

to invite prominent economists, philosophers, historians, policy-makers and social scientists in general who have had some association with Sen—as student, colleague, editor, co-author or co-activist—to contribute to what would be a major multi-volume work. Yet so many people have been associated with Amartya Sen in one way or the other and so many have been profoundly influenced by him that there was no way we could accommodate them all. We are aware that this is an idiosyncratic collection of authors, and the list of those who could legitimately make claims to being a part of this project, but had to be left out for reasons of space, is long.

Amartya Sen did not start out in welfare economics. As an undergraduate student at Trinity College, Cambridge University (where he later returned as Master), Sen wrote a dissertation on the choice of technology. That dissertation gained him a Ph.D. and in 1960 was published as a book. *Choice of Techniques* became a celebrated text, which students of development, planning and growth read, did research on and modified. This work had many spin-offs for Sen himself. He contributed to growth theory, cost–benefit analysis and development; he became full professor at the Delhi School of Economics in 1963.

During this time his interests were, however, beginning to shift as he became intrigued by the problems of welfare and the conundrums of voting and social choice. There was a small literature on the latter at that time. It had all begun in the early fifties with the publication of Kenneth Arrow's path-breaking "impossibility theorem". Amartya began working on this seriously after he moved to the Delhi School of Economics and this resulted in his seminal book *Collective Choice and Social Welfare*. With a diverse set of ideas, from mathematical logic through welfare economics to ethics, crowding its pages, the book quickly became a classic. It drew the attention of not just economists, but also professional philosophers.

In the 1970s Amartya Sen moved to England, first as Professor of Economics at the London School of Economics (LSE) and later to Oxford, where he became Drummond Professor. It is at this stage that we, the editors of this book, got to know him. Kaushik Basu did his Ph.D. with him from 1974 to 1976 at the LSE. Ravi Kanbur had him as his D.Phil. adviser at Oxford from 1978 to 1980.

During his stay at the LSE, Amartya Sen's interest moved on to something which had always been there, though somewhat subliminally—the subject of poverty and inequality. He published his celebrated book, *The Economics of Inequality*, and later the seminal paper on poverty measurement in *Econometrica*. After seven years at the LSE, he moved to Oxford in 1977. And his interest in philosophy flourished. The first evidence of Amartya's interest in philosophy was a paper he wrote in 1959 in *Enquiry*. This little-known paper consisted of a lucid elaboration of the conflict between determinism and free will. Sen's philosophical interest continued to grow in Oxford, where he interacted with the philosophers, and his papers appeared in *Philosophical Quarterly*, *Mind*, *Philosophy and Public Affairs*, *Theory and Decision* and the *Journal of Philosophy*. When he moved to Harvard in 1987 his appointment

would be in the economics and philosophy departments and he offered a course jointly with Robert Nozick.

His work on poverty and basic needs, coupled with his interest in moral philosophy, led Sen to question the pervasive use of income to measure and compare human welfare. Up until then, the most popular way of ranking the economic performance of countries had been in terms of the per capita incomes of the people. Sen argued repeatedly for the need to bring in other indicators of the standard of living in order to evaluate the overall well-being of societies. This work has had enormous practical implications, ever since the United Nations Development Programme, openly acknowledging Sen's influence, began computing the Human Development Index (HDI) to evaluate societies.

Since moving to Harvard in the 1980s, and during his term as Master of Trinity College, Cambridge in the 1990s before returning to Harvard, Sen's academic reach and popular engagement has continued to broaden. He has written on identity, on history and on culture. He has engaged vigorously in policy debates, advancing an egalitarian, tolerant and pluralistic perspective on the issues of the day. Sen's later work has come under criticism from academic writers for being populist but that is more a reflection on his critics than on him. A large part of his recent work is that of a person with a cause, and he has devoted his considerable analytical power to trying to win over the hearts and minds of the ordinary person, journalists and policy-makers and not simply engaging in debates with other academics. This argumentative Indian, to use the title of one of his recent books, has joined and enriched the arguments for a better world.

It is this full range of Sen that we try to engage with in this rather mammoth project. Our two volumes mirror Amartya Sen's own two-volume collected works—*Choice, Welfare and Measurement* and *Resources, Values and Development*—and, in addition, we try to stretch the contours further to accommodate his expanding range of interests since these collected works were published in 1982 and 1984 respectively. We have called our two volumes, respectively, *Ethics, Welfare and Measurement* and *Society, Institutions and Development*.

The first volume begins with matters of ethics and moral philosophy—appropriately we feel, given the foundational status of philosophy and given that so much of modern social choice and welfare draws on principles of moral philosophy and the philosophy of human rights. Several essays in this collection then go into matters of social choice theory and evaluative measurement, matters that were close to Amartya Sen's heart through the seventies and eighties. His influential papers and books on the "capabilities approach" to social evaluation appeared mostly after the publication of his collected works, and this has been enormously influential in the social sciences, perhaps even more outside of economics than within. We felt that this deserves a full subsection and so there is a cluster of eight papers on the subject. The first volume ends with some more applied essays on collective action and also some on Sen's recent interest and an urgent subject of our times—identity.

Volume II is on the subject of development, on which Amartya Sen has contributed extensively throughout his life, and which, we have noted, was widely (and wrongly) believed to be the work for which he was awarded the Nobel Prize. But we decided to place this subject within a much larger social science context than usual. Thus this second volume has contributions from political scientists, sociologists, philosophers and historians, and of course there is the ubiquitous presence of economists. In its first part the second volume picks up from where the first volume left off—the subject of capabilities in actual developmental settings. The second part turns to the topic of gender and the household. Sen is a prominent figure in the field of feminist economics and his contributions on differences in the treatment of the girl child and boy child have been of great influence for both academic research and policy-making. His writing on the millions of "missing girls" in India and China stirred consciences and drew the attention of bureaucrats and politicians. He has also worked, in more academic ways, on the problem of decision-making within the household by applying ideas of Nash bargaining to the nature of "cooperative conflict" among householders. These interests are reflected in the papers in Part II.

Interestingly, among Amartya Sen's earliest research were several papers on growth theory and a famous edited volume on the subject of growth. The third part of Volume II consists of papers on growth but with special attention to the urgent problems of poverty, climate change and labor standards. The final part of the volume deals with the role of sociology, politics and history in the context of development and economics. While one may debate the end *of* history as intellectuals have done in recent times, we for our part wanted to emphasize the importance of this subject for economics by ending *with* history. Scholars have argued that we need to embed economics in the related social sciences to enrich our understanding of not just the social sciences, but economics itself. The last eight essays try to carry this agenda forward, however slowly.

This sizeable project would not have been possible but for the supportive atmosphere of Cornell University. Our faculty colleagues have been a source of discussion and stimulus, so important for this kind of work. Kiran Gajwani helped greatly with research assistance. Sue Snyder provided excellent administrative support throughout in keeping the volumes on track. Sarah Caro of Oxford University Press has been an enthusiastic backer at each of the many long stages of this project, from inception to production. We thank them all.

Ithaca, New York
4 March 2008

PART I

HUMAN DEVELOPMENT AND CAPABILITIES

This page intentionally left blank

CHAPTER 1

INTER-COUNTRY COMPARISONS OF INCOME POVERTY BASED ON A CAPABILITY APPROACH

SANJAY G. REDDY
SUJATA VISARIA
MUHAMMAD ASALI

I. INTRODUCTION

How should poverty be estimated? Amartya Sen has argued persuasively that poverty must be seen as the deprivation of basic capabilities, where capabilities are the "substantive freedoms [a person] enjoys to lead the kind of life he or she has

We are grateful to Carlos Sobrado of the World Bank, Emilian Karugendo of the National Bureau of Statistics of Tanzania and Patrick Ward of Oxford Policy Management, and the General Statistics Office of Vietnam for their help with household surveys for Nicaragua, Tanzania and Vietnam

reason to value", rather than merely as low income (Sen 1999: 87). Income is one instrument for attaining such substantive freedoms, but only one. Moreover, "the instrumental relation between low income and low capability is *variable* between different communities and even between different families and different individuals (the impact of income on capabilities is contingent and conditional)" (Sen 1999: 88). Sen has also pointed out that, more generally, all poverty assessment involves two component exercises: the identification of the poor (i.e. the determination of who is poor and to what extent) and the aggregation of this information to form a judgement concerning the extent of poverty in a society. An identification criterion that is uniform at *some* level of abstraction must be applied to all individuals if this exercise is to provide a meaningful basis for comparisons. For example, we may define as income-poor all those whose money income falls below a certain level, or instead we might define as income-poor all those whose money income is below the level required to achieve some end (such as the attainment of basic capabilities, as Sen recommends). For an exercise of poverty assessment to be meaningful, it is necessary (although not sufficient) that it be possible to view it as consistently applying a uniform identification criterion to all individuals.

Efforts to assess poverty at the regional and global levels are as subject to this demand as are poverty assessments within the national context. Meaningful inter-country income comparison and aggregation requires that a common identification criterion be applied in all countries. The predominant method in use at present for such comparison and aggregation is the "money-metric" approach. In this approach, the identification criterion depends on an international poverty line (IPL) expressed in PPP dollars of a specific year and converted into poverty lines expressed in local currency units (and deemed equivalent to the IPL). Although it may appear that this approach establishes a uniform identification criterion, it may do so only in an ultimately futile sense. As argued by Reddy and Pogge (forthcoming), the PPP conversion factors used for this purpose do not reflect an invariant level of purchasing power over essential commodities. Therefore, the substantive interpretation of existing $1 and $2 per day IPLs is widely discrepant across countries. If it is demanded that the identification criterion have a common substantive interpretation in order to be deemed uniform in the relevant sense,

respectively. We thank Shaohua Chen for answering several questions about the World Bank's international poverty lines in these three countries. Howard Nye played a role in conceptualizing and implementing this project at an early stage. The Bureau of Development Policy and the International Poverty Centre of the United Nations Development Programme provided essential support. We thank Terry McKinley for facilitating this support. We are also grateful to Ravi Kanbur, Camelia Minoiu, Nanak Kakwani, Terry McKinley, participants in the international conference on the "many dimensions of poverty" held by the UNDP International Poverty Centre, and participants in a workshop on capabilities and well-being evaluation held at the University of California at Riverside for helpful comments. Our greatest thanks are to Amartya Sen for the abiding inspiration that he has provided.

then the existing money-metric approach does not provide the uniform identification criterion that is required (and that would be achieved if it reliably referred to an invariant level of purchasing power over essential commodities).

There is another problem with the money-metric approach. A meaningful poverty line should, at a minimum, reflect the cost of achieving basic human requirements. Although there can be reasonable disagreement about how to understand such requirements, there cannot plausibly be disagreement that a persuasive poverty line should reflect them. A poverty line is meaningful only if persons with incomes falling below the poverty line can reasonably be thought of as poor. Unfortunately, the local currency "equivalents" of the IPL not only have discrepant interpretations across countries but often fail to reflect the cost of achieving basic human requirements in each individual country. It is difficult to make the case that the "$1/day" and "$2/day" international poverty lines reflect the cost of achieving the real requirements of human beings (see e.g. Reddy and Pogge forthcoming). The poverty lines employed in an income poverty assessment must be evaluatively meaningful if the exercise of income poverty assessment as a whole is to be meaningful, but this test is failed by the money-metric approach.

A fully meaningful approach to inter-country income poverty comparison and aggregation would establish a poverty line for each country (or perhaps subnational jurisdiction) corresponding to the minimum cost (in that country) of achieving a certain set of basic human requirements (or as we prefer to understand them, income-dependent elementary human capabilities) uniformly conceived across countries. The same elementary human capabilities would be used to define the poverty line in each country. The resulting poverty lines would embody a uniform identification criterion possessing the advantage of having the same meaningful interpretation in all countries. Such an approach would avoid using PPPs altogether, and eliminate both problems of the current money-metric approach in a single stroke. Conceptually, the capability-based alternative approach to income poverty assessment involves nothing more than the generalization to the context of inter-country comparisons of an approach that is already widely used and thought of as appropriate at the national level. Of course, a capability-based approach to income poverty assessment is not the same as an approach to poverty assessment that assesses capability deprivations as such. There is room and necessity for both approaches to poverty assessment. It may surprise advocates of the capability approach to hear methods which they have viewed as inadequate in the national context described here as "capability-based" but they do merit this description in contrast to the "money-metric" approach.

We report here the results of a capability-based approach to income poverty assessment. We show that it is possible to use existing household survey data from three different countries (Nicaragua, Tanzania and Vietnam) on three continents to define a uniform capability-based criterion for identifying the poor. We focus

centrally on the capability to be adequately nourished, as it is both universally agreed to be a relevant basic capability and is easy to employ. We use this criterion to establish poverty lines that possess a common capability-based interpretation (in terms of nutritional non-deprivation) in all three countries and then estimate income poverty in these countries. By definition, the resulting estimates are comparable in the sense that they refer to the same (capability-based) concept of poverty in all three countries. We thus demonstrate that, even with existing data sources (which have not been specifically designed with the purpose of supporting such comparisons), it is possible to implement a capability-based approach to global income poverty estimation. The sense in which the approach to poverty assessment adopted here is capability-based is admittedly a very limited one. It focuses on explicitly specifying a single capability (the ability to be adequately nourished) while making indirect allowance for other relevant capabilities. It also takes a rather restricted approach (based on food energy requirements) to the empirical identification of that capability. Finally, no allowance is made for variations in the commodities required for achieving basic capabilities, as is ultimately required in a comprehensive capability-based perspective. The approach pursued therefore falls far short of the best possible. Nevertheless, it presents a superior alternative to the money-metric approach, in that it is in contrast grounded in a conception of basic human requirements, and employs this conception uniformly across countries. For this reason, although we are aware of the limitations of this characterization, we will refer to the approach as "capability-based". It is obvious that various enhancements can and could be undertaken to generate more fully adequate income poverty assessments for each country (for example, through using household adult-equivalence scales). However, the desirability of undertaking such enhancements applies in common to all approaches to regional and global income poverty estimation.[1]

We contrast the poverty estimates that we obtain on the basis of capability-based poverty lines with those based on the money-metric international poverty lines that are commonly used and show that our approach yields notably different results. We also examine how the use of capability-based poverty lines, instead of money-metric IPLs, affects cardinal and ordinal comparisons of poverty across countries and over time. On the basis of this exercise, we argue that there is no "quick fix" that can align the existing money-metric poverty lines with a capability-based concept of poverty. A simple increase or decrease in the money-metric IPL without a change in the PPPs used to convert the IPL into local currency units cannot bring about such alignment, because the adjustment that is required varies from country to country. A more comprehensive program of capability-based poverty

[1] Notably, existing global poverty estimates based on money-metric IPLs produced by the World Bank and others have not employed household equivalence scales.

line construction (and complementary survey design) offers the best way forward for inter-country poverty comparison and aggregation.

The poverty estimates produced here are not authoritative estimates of poverty in each country, since the data sources and the methods of poverty line construction applied here are insufficiently refined to support the claim that the estimates are definitive. Our method of arriving at the poverty line is but one of several possible non-money-metric methods for constructing a set of poverty lines that possesses a more uniform and meaningful interpretation across countries.

We have taken the methodology for poverty line construction used in the 1993 Vietnam Living Standards Measurement Survey (LSMS) as our starting point. We may infer from its adoption that the method was considered acceptable for measuring national poverty in Vietnam. This starting point is, to a degree, arbitrary. It represents one among many plausible ways of constructing a nutritionally anchored poverty line (see e.g. Ravallion 1994). We also apply this methodology of poverty line construction to Tanzania and Nicaragua. Finally, we compare the resulting estimates with existing national poverty estimates for Tanzania and Nicaragua, and also with those from the money-metric IPL approach.

We find that the choice of approach matters a great deal. In comparing income poverty estimates across countries and over time, the capability-based approach that we employ does, in some instances, give results significantly different to those of the money-metric approach. Both cardinal comparisons and (perhaps more surprisingly) ordinal rankings of income poverty across countries are influenced by the approach used. The rest of the chapter is organized as follows. In the next section, we describe the conceptual content of the method we apply. In section III, we describe the methodology used in each country and in section IV we describe the resulting poverty estimates. Section V discusses the implications of our analysis for inter-country poverty comparison and aggregation and presents our conclusions.

II. Inter-Country Income Poverty Comparison and Aggregation Using Existing Data: A Rough Method

The first step in the exercise is to identify a relevant set of elementary capabilities. The cost of achieving these elementary capabilities can be described in a familiar manner. It is assumed that for each individual there exists some set of commodity bundles which suffices to achieve the elementary capabilities. Given the prices faced

by an individual, we can identify the minimum cost of achieving the elementary capabilities.

In a particularly simple approach, the adequacy set is assumed to be common for all persons. We follow the 1993 Vietnam LSMS in adopting this approach. As noted, this approach is a mere starting point, and insufficiently attentive to the diverse features of persons (e.g. age, gender, occupation) which influence the way in which they can transform commodities into capabilities. These diversities should be taken into account in a fully adequate approach to poverty assessment.[2]

In our empirical exercise, we take the ability to be adequately nourished as the centrally relevant elementary capability which anchors the identification exercise, at the risk of considerable over-simplification. If it is assumed that a certain fixed level of calories is sufficient for all persons to achieve adequate nourishment, then the minimum cost of achieving this capability may be identified for all persons. In this paper, we operationalize this idea in a particular way, following the Vietnam LSMS. We choose as a reference group that quintile of the population which comes closest to achieving the nutritional standard (in our case, a food-energy standard—2,100 kilocalories (kcals)). For simplicity, the consumption pattern of this reference group is taken to indicate the composition of the minimum-cost bundle. The food poverty line is the cost of the bundle containing exactly 2,100 kcals and reflecting this consumption pattern. This method takes into account the preferred patterns of food consumption of the group in the population whose consumption is closest to the nutritional standard, and is a rough and ready way of making allowance for prevailing consumption norms. Although there is the danger of using a consumption pattern that is "richer" in one country than in another as if they were equivalent, there seems no straightforward way to avoid this problem without bringing in auxiliary judgements. These could (and indeed should) be integrated into more comprehensive exercisese.

Next, we make an allowance for non-food requirements. We determine the ratio of non-food to food expenditure for the reference population and then maintain this ratio at the poverty line. This is a highly inadequate approach to a complex problem; we follow it here because of a lack of independent information on non-food requirements and costs.

Suppose that the average commodity bundle of the reference population has a calorie content that falls below 2,100 kcals by x per cent. Our approach assumes that the reference population's shortfall in the expenditure necessary to achieve both the food and the non-food expenditure requirements (for capability adequacy) is also x per cent. Let us call this the equiproportionality assumption. The implied food and non-food poverty lines are added to constitute a general poverty line which is assumed to reflect the minimum cost of achieving non-poverty.

[2] In many national poverty estimation exercises, this problem is imperfectly addressed with the use of equivalence scales.

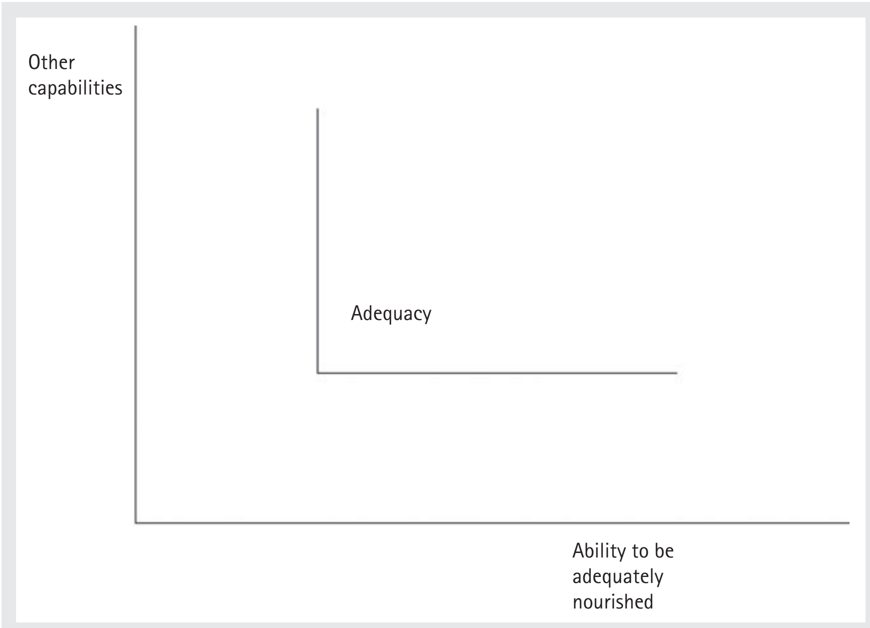


Fig. 1.1. Capability space

We may informally illustrate our general approach as follows, assuming indexable capabilities. If a minimally adequate level of each of the relevant income-dependent capabilities is deemed essential to avoid income poverty, this gives rise to an "adequacy set" with an L-shaped lower contour in the capability space (see Fig. 1.1).

We next translate this concept of poverty into terms which are more amenable to measurement. An adequately nourished individual needs to receive adequate amounts of various food characteristics: food energy, protein, fats, fiber, micronutrients and so on.[3] It may be thought appropriate to make allowance for adequate amounts of other commodity characteristics as well (e.g. taste). Since different commodities contain these characteristics in different proportions, substitution between them may be possible, giving rise to a differently shaped (even smooth) lower contour of the adequacy set as represented in characteristics space (see Fig. 1.2). For example, it is conceivable that a lower level of food energy intake may suffice for nutritional adequacy if fat, protein, fiber or other nutrients are contained in the diet to a greater extent. Tradeoffs of this type may be pervasive. However, our approach will not take explicit note of this possibility. For simplicity, researchers have focused historically on the food energy intake of individuals and have anchored the poverty line in a calorie adequacy threshold. We will not depart from this classical approach, despite its very severe limitations,

[3] On the concept of characteristics of commodities, see Lancaster (1971).

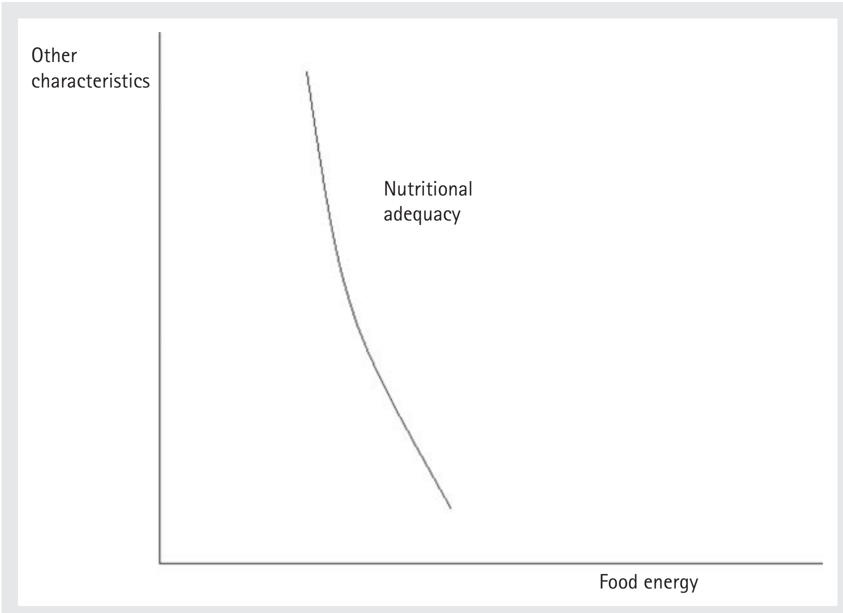


Fig. 1.2. Characteristics space

as employing it will suffice for us to make our broader methodological point which is concerned with bringing to light the limitations of the money-metric approach. In our study, the calorie adequacy threshold is defined as 2,100 kcals per day.

In principle, it should be possible to relax the equiproportionality assumption. However, in the absence of consensus on what non-food capabilities are of concern, on the characteristics of the commodities which promote them, on the transformation function that relates these characteristics to capabilities, and on the levels of each capability that ought to be deemed minimally adequate, any adjustment requires justification. There is a need for more explicit specification of the non-food requirements and the collection of information required to determine the cost of meeting these requirements. Such an exercise will not be feasible without the design of surveys specifically with this end in mind, and complementary exercises in evaluative judgement.

III. Data and Empirical Work

The methodology described in the previous section is applied to three countries: Nicaragua, Tanzania and Vietnam. The most important feature of the exercise is

Table 1.1. World Bank poverty headcount ratio estimates

	1991		1993		1998	
	$1/day	$2/day	$1/day	$2/day	$1/day	$2/day
Nicaragua	—	—	47.94	77.78	44.71	79.03
Tanzania	48.54	72.53	—	—	—	—
Vietnam	—	—	14.63	58.16	3.8	39.68

Source: World Bank's World Development Indicators (accessed online 13 March 2005).

that we use a *common* capability-based approach in all three countries. We use these poverty lines to compute poverty estimates, and then compare them to those from money-metric $1 per day and $2 per day international poverty lines. We then explore the robustness of inter-country poverty comparisons and aggregates with respect to the choice of identification concept. Despite the necessarily second-best nature of the exercise, we believe that it represents a more coherent and meaningful approach for inter-country comparisons of poverty than does the prevalent money-metric approach.[4]

The countries selected for this exercise are attractive choices for a few distinct reasons. First, each country lies in a different continent, thus allowing us to demonstrate that a limited type of capability-based inter-country comparison and aggregation of poverty estimates can be undertaken despite different food habits and non-food expenditure patterns. Second, two of the countries (Nicaragua and Tanzania) had very similar headcount ratios in the 1990s according to the World Bank's estimates based on its $1 and $2 per day IPLs, but the third country (Vietnam) had a very different headcount ratio, summarized in Table 1.1.

We also compute standard errors of all poverty measures by using bootstrapping. Thus we can make both ordinal and cardinal comparisons across countries and over years and check if the differences are statistically significant.

Third, in each of these countries, there are well-designed household surveys to which we could gain access. For Vietnam and Nicaragua, the data are from the Living Standard Measurement Surveys conducted in these countries by the World Bank in collaboration with national statistical agencies. The data on Tanzania come from the Household Budget Survey conducted by the Tanzanian National Bureau of Statistics.

In order to facilitate comparison of statistics across countries and across poverty line concepts, we also calculated bootstrapped standard errors (using 1,000 iterations) for every poverty estimate. The large number of iterations guaranteed a very

[4] The poverty estimates produced by the Economic Commission for Latin America and the Caribbean (ECLAC) are an important exception to the dominant use of the money-metric approach (Altimir 1982).

high confidence level in most, if not all, cases in the calculation of the standard errors: a 5 per cent significance level and a deviation in magnitude of approximately 4.5 per cent from the limiting standard deviation.[5]

III.1 Methodology used for Vietnam

The methodology applied in Vietnam amounts to five steps:

1. Exogenously identify a threshold of nutritional capability adequacy and characterize it in terms of characteristics of commodities consumed (the 2,100-kcal norm).
2. Identify the quintile whose average calorie intake is closest to the calorie threshold.
3. Determine the cost of achieving this threshold (the food poverty line) while maintaining the pattern of consumption of a reference quintile.
4. Establish an allowance for non-food expenditures such that the ratio of this allowance to the food poverty line is the same as the ratio of non-food to food expenditures for the reference quintile.
5. Set an overall poverty line, equal to the sum of the food poverty line and the non-food expenditure allowance, and determine the number of persons living in households with per capita consumption beneath this level.

The estimates of poverty for Vietnam according to various poverty indicators are reported in Table 1.4.

III.2 Applying the Methodology to Nicaraguan Data

The data for Nicaragua are from the Nicaraguan LSMS for 1997/8 (known as the EMNV 1998 Survey). We have followed the methodology used in Vietnam to calculate the capability-based poverty line for Nicaragua, employing both general and food-based CPIs to calculate equivalents over time.

Once we had computed the poverty line for Nicaragua, the next step involved calculating income poverty estimates. From the household-level data set, we created an expanded individual-level data set in which each member of each household

[5] We used the method proposed in Andrews and Buchinsky (2000) to choose the optimal number of bootstrap iterations, and to evaluate the performance and precision of the resulting bootstrapped standard errors. In fact, following the procedures proposed by Deaton (1997) and Howes and Lanjouw (1998), we calculate standard errors both using bootstrapping and using the sepov command in STATA. The latter implements a standard error calculation based on theoretical premises. In both instances, a simple two-stage sampling design is assumed, whereas in fact all of the surveys we have examined involve a more complicated survey design. As a result, the standard errors we calculate cannot be viewed as more than indicative. We report and refer only to the bootstrapped standard errors, since the standard errors calculated through the two approaches were generally very close.

Table 1.2. Annual poverty lines, Nicaraguan cordobas, 1998

$1/day general CPI	4,017.20
$2/day general CPI	8,034.40
$1/day food CPI	4,119.44
$2/day food CPI	8,238.87
Capability-based	3,018.42

was assigned the annual per capita expenditure of that household. We then calculated the headcount ratio: the proportion of persons in the population whose per capita expenditure was below the poverty line. Similarly we computed the aggregate poverty gap, income gap ratio, Sen Index and the Foster-Greer-Thorbecke Indices, with values of α equal to 1.5, 2, 2.5, 3, 3.5, and 4, and calculated standard errors (the methodology is discussed further below) so as to judge the precision with which the poverty measures were estimated.

Next, we compared our capability-based estimates of income poverty in Nicaragua with the estimates that the money-metric methodology would have produced. The comparison was done with the poverty estimates corresponding to different poverty lines: the $1 PPP per day and $2 PPP per day poverty lines adjusted by the consumer price index or a food price index for the country.[6] The poverty lines are presented in Table 1.2.

The table implies that the capability-based estimates are lower than the $1 per day estimates. This can be confirmed by referring to Table 1.5, which reports Nicaragua income poverty estimates for varying poverty lines and measures of poverty.

III.3 Applying the Methodology to Tanzanian Data

The data for Tanzania are from the 2000/1 Tanzanian Household Budget Survey (HBS), conducted by the National Bureau of Statistics between May 2000 and June 2001. Once again, we applied our chosen methodology to establish a poverty line for Tanzania.

We produced income poverty estimates based on our capability-based poverty line for Tanzania. We provide a summary of the results based on our capability-based income poverty line and on the $1 and $2 PPP per day income poverty lines. Once again, we used both the general CPI and a food CPI to convert the

[6] Shaohua Chen of the World Bank kindly provided us with the consumer price indices. These originate in the World Bank's Development Data Group and are the same ones used in the Bank's global poverty assessments. The food price indices we used are produced by the ILO and available in the World Bank's World Development Indicators.

Table 1.3. Annual poverty lines, Tanzanian shillings, 2000/2001

$1/day general CPI	147,613.5
$2/day general CPI	295,227
$1/day food CPI	158,410.83
$2/day food CPI	316,821.66
Capability-based	80,365.1

IPL from local currency units in the base year to the local currency units of the survey year. Since the Household Budget Survey was administered over the period of a whole year from mid-2000 to mid-2001, we used the geometric means of the price indices pertaining to the relevant years. We report the different poverty lines that we employed in Table 1.3.

Our detailed poverty estimates for different poverty lines and measures of poverty for Tanzania are presented in Table 1.6.

IV. Inter-Country Income Poverty Comparison and Aggregation According to Alternative Approaches: Results

Tables 1.4–1.6 present the three types of poverty estimates for the different country-years: Vietnam in 1993 and 1998, Nicaragua in 1998 and Tanzania in 2000/1. The results are based on three different poverty lines: the $1 a day, $2 a day, and the capability-based poverty lines. Both the $1 a day and $2 a day money-metric poverty lines employed are those defined by the World Bank for a particular base year, 1993. As noted, we use both general and food price indices to adjust these poverty lines to their assessment year equivalents.

In the tables, the type of poverty line used is described in the first row. We provide estimates for the headcount ratio, income gap ratio, and poverty gap ratio, along with the aggregate poverty gap, Sen Index and the Foster–Greer–Thorbeck Indices for different values of its distribution sensitivity parameter. For each poverty estimate, the associated bootstrapped standard error is in parentheses.

We ask three kinds of questions:

1. Does the extent of estimated poverty in each country depend on the poverty identification concept?
2. Do the ordinal and cardinal comparisons among country-years depend on the poverty identification concept?
3. Does the poverty identification concept influence the estimated extent of aggregate poverty and the share of that aggregate in different countries?

To examine the first question, consider initially the case of Tanzania in 2000/1 (Table 1.6). Columns (1) and (3) report estimates based on a $1 a day poverty line, using the food CPI and the general CPI respectively. Columns (2) and (4) report estimates for the $2 a day poverty line. Column (5) reports the poverty estimates for the capability-based poverty line. Each row corresponds to a different poverty measure. We can see that the capability-based poverty line consistently gives lower estimates than the estimates based on $1 a day, regardless of the poverty measure used.

The reduction is substantial. Whereas according to the $1 a day poverty line, 75 per cent of the Tanzanian population is poor, according to the capability-based poverty line, only 40 per cent is poor. A similar pattern can be seen in the results for Nicaragua as well (Table 1.5), although the reductions are less drastic. Whereas the use of the $1 a day poverty line generates a 44.6 per cent headcount ratio, the headcount ratio associated with the capability-based poverty line is 30.61 per cent.

On the other hand, for Vietnam in 1993, the use of the capability-based poverty line gives rise to much higher poverty estimates than the $1 a day poverty line, although they are below the $2 a day estimates. This is true for Vietnam in 1998 as well. The presence of data for two different years for Vietnam also allows one to see if the choice of poverty line affects the rate of poverty reduction. According to the $1 a day poverty line, the headcount ratio fell from 13.4 per cent in 1993 to 5.2 per cent in 1998, a reduction of 61 per cent. According to the $2 a day poverty line, the reduction was 34 per cent. Once again, the use of the capability-based poverty line gives rise to a rate of reduction that is between the two, at 38 per cent (see Table 1.7).

However imperfect our capability-based approach might be, it was constructed with the explicit aim of capturing the minimum cost of achieving the same basic capabilities in *each* of these three countries. In light of this, the fact that our estimates differ drastically from the money-metric estimates is informative. It raises the concern that the money-metric poverty lines fail to represent the cost of achieving basic capabilities in these countries.

In answer to the second question, we find that the ordinal rankings of country-years according to the extent of poverty are often robust with regard to the choice of identification concept. In Table 1.8, dominance relations are represented in a Hasse diagram (following the suggestion made by Amartya Sen in diverse writings that

Table 1.4. Income poverty statistics, Vietnam, 1993–1998

	Poverty line					
	1993			1998		
	$1/day	$2/day	Capability-based	$1/day	$2/day	Capability-based
HCR	13.37 (1.270)	63.72 (1.750)	58.15 (1.785)	5.20 (.710)	41.98 (1.626)	35.62 (1.672)
IGR	21.12 (1.729)	34.22 (.846)	31.78 (.853)	17.15 (1.546)	27.13 (.915)	25.43 (.923)
APG (m)	0.42 (.065)	6.54 (.297)	5.11 (.258)	644.80 (121)	16,470 (1,150)	12,070 (950)
PGR	2.82 (.433)	21.81 (.953)	18.48 (.905)	0.89 (.166)	11.39 (.734)	9.06 (.669)
Sen	4.04 (.625)	28.67 (1.201)	24.64 (1.169)	1.30 (.237)	15.56 (.951)	12.50 (.880)
FGT (1.5)	1.59 (.305)	14.25 (.751)	11.79 (.698)	0.46 (.099)	6.87 (.521)	5.34 (.458)
FGT (2)	0.98 (.228)	9.72 (.606)	7.88 (.554)	0.26 (.062)	4.38 (.378)	3.34 (.323)
FGT (2.5)	0.64 (.177)	6.85 (.498)	5.45 (.450)	0.16 (.041)	2.91 (.280)	2.19 (.234)
FGT (3)	0.44 (.141)	4.95 (.416)	3.89 (.373)	0.10 (.028)	2.00 (.212)	1.48 (.174)
FGT (3.5)	0.32 (.115)	3.66 (.352)	2.84 (.314)	0.07 (.019)	1.41 (.163)	1.03 (.132)
FGT (4)	0.24 (.095)	2.76 (.302)	2.12 (.268)	0.04 (.013)	1.02 (.127)	0.74 (.102)

Notes: Bootstrapped standard errors are in parentheses; see text for details. The $1 a day poverty line for 1993 is 629,341.1 dongs; the capability-based poverty line for 1993 is 1,160,363 dongs. The $1 a day poverty line for 1998 is 953,794 dongs; the capability-based poverty line for 1998 is 1,758,581 dongs.

Table 1.5. Income poverty statistics, Nicaragua, 1998

	Poverty line				
	$1 food-CPI	$2 food-CPI	$1 general-CPI	$2 general-CPI	Capability-based
HCR	45.78 (1.310)	79.90 (1.229)	44.62 (1.310)	79.03 (1.265)	30.61 (1.464)
IGR	37.80 (.934)	52.43 (.665)	37.19 (.976)	51.80 (.678)	31.66 (.836)
APG (m)	3432 (154)	16,620 (607)	3209 (146)	15,830 (581)	1409 (79.800)
PGR	17.30 (.720)	41.89 (.840)	16.59 (.712)	40.93 (.837)	9.69 (.558)
Sen	22.98 (.875)	52.21 (.951)	22.12 (.862)	51.17 (.959)	13.25 (.741)
FGT (1.5)	11.99 (.573)	32.60 (.757)	11.44 (.562)	31.73 (.753)	6.31 (.401)
FGT (2)	8.67 (.461)	26.06 (.691)	8.24 (.448)	25.27 (.686)	4.33 (.301)
FGT (2.5)	6.46 (.374)	21.25 (.634)	6.12 (.362)	20.54 (.627)	3.09 (.232)
FGT (3)	4.93 (.307)	17.59 (.581)	4.66 (.296)	16.96 (.574)	2.26 (.183)
FGT (3.5)	3.84 (2.56)	14.74 (.533)	3.61 (.245)	14.18 (.524)	1.70 (.147)
FGT (4)	3.04 (.214)	12.49 (.488)	2.85 (.205)	11.98 (.479)	1.30 (.119)

Notes: Bootstrapped standard errors in parentheses; see text for details. The $1 a day food-CPI poverty line is 4,119.437 cordobas; the $1 a day general-CPI poverty line is 4,017.20 cordobas; the capability-based poverty line is 3,018.42 cordobas.

Table 1.6. Income poverty statistics, Tanzania, 2000/2001

	Poverty line				
	$1 food-CPI (1)	$2 food-CPI (2)	$1 general-CPI (3)	$2 general-CPI (4)	Capability-based (5)
HCR	78.51 (1.218)	95.66 (.390)	75.39 (1.321)	94.75 (.518)	40.13 (1.756)
IGR	47.84 (.850)	66.60 (.678)	45.99 (.858)	64.80 (.698)	31.45 (1.092)
APG (m)	1,898,000 (110,000)	6,438,000 (313,000)	1,632,000 (97,400)	5,782,000 (285,000)	323,500 (27,800)
PGR	37.56 (1.076)	63.70 (.803)	34.67 (1.077)	61.40 (.838)	12.62 (.835)
Sen	47.21 (1.204)	73.64 (.713)	43.91 (1.233)	71.55 (.781)	17.25 (1.069)
FGT (1.5)	28.07 (.970)	53.78 (.872)	25.53 (.953)	51.30 (.897)	8.19 (.624)
FGT (2)	21.59 (.866)	45.99 (.899)	19.39 (.838)	43.47 (.913)	5.60 (.474)
FGT (2.5)	16.98 (.770)	39.72 (.900)	15.07 (.736)	37.24 (.904)	3.98 (.365)
FGT (3)	13.58 (.684)	34.59 (.885)	11.94 (.646)	32.18 (.881)	2.91 (.285)
FGT (3.5)	11.02 (.607)	30.32 (.860)	9.60 (.567)	28.02 (.849)	2.17 (.224)
FGT (4)	9.04 (.539)	26.73 (.829)	7.82 (.499)	24.55 (.813)	1.65 (.179)

Notes: Bootstrapped standard errors in parentheses; see text for details. The $1 a day food-CPI poverty line is 158,410.83 Tanzanian shillings (TSH); the $1 a day general-CPI poverty line is 147,613.5 TSH; the capability-based poverty line is 80,365.10 TSH.

Table 1.7. Vietnam headcount ratio (HCR) reduction, 1993–1998

	1993 HCR	1998 HCR	1998 HCR/1993 HCR
$1/day	13%	5%	0.38462
$2/day	64%	42%	0.65625
Capability-based	58%	36%	0.62069

intersection partial orderings can be a valuable device in empirical investigations). A dominance relation is identified to exist only if one measure can be deemed greater than another at the 95 per cent level of confidence. The dominance relations are represented by a vertical hierarchy: country-years with greater poverty are placed in a tier vertically above country-years with less poverty. Countries which do not stand in any dominance relation to one another are placed in the same tier. For example, consider the capability-based estimates of the headcount ratio (HCR). The diagram shows that Vietnam in 1993 had a higher HCR than Vietnam in 1998, at a 95 per cent significance level. It was also higher than Tanzania 2000/1, which in turn, together with Vietnam 1998, was higher than Nicaragua 1998. The HCRs of Tanzania 2000/1 and Vietnam 1998 are not significantly different from each other. It can also be seen that Tanzania 2000/1 is estimated to have had greater poverty than Nicaragua is estimated to have had in 1998 for almost all the measures. Similarly, Vietnam in 1993 is estimated to have had greater poverty than Vietnam in 1998 for almost all the measures. Thus, some dominance relations are stable, irrespective of the concept underpinning the poverty line or the poverty measure used.

However, some dominance relations are altered drastically. The money-metric IPL-based poverty estimates almost always suggest that poverty was greatest in Tanzania 2000/1, second greatest in Nicaragua 1998, third greatest in Vietnam in 1993 and lowest in Vietnam 1998. In sharp contrast, the capability-based estimates suggest that poverty was almost always highest in Vietnam in 1993. However, it is ambiguous whether it was lowest in Vietnam 1998 or in Nicaragua 1998.

An important observation emerges from this table. Income poverty appears to have decreased in Vietnam from 1993 to 1998, regardless of the method used. There is a broad-based perception that there was a large decrease in poverty in Vietnam in the 1990s. It is hence reassuring that the capability-based results confirm this. This reduction is apparent in the money-metric estimates as well. However, when we compare countries (for example, Tanzania 2000/1 with Vietnam 1993), the direction of ordinal comparisons depends on the choice of the poverty identification concept. It may be confirmed that the ordinal comparisons between country-years are almost uniformly invariant with regard to the choice between money-metric ($1 or $2 per day) IPLs. On the other hand, ordinal comparisons between country-years are greatly influenced by the choice between a capability-based and

Table 1.8. Hasse diagram for Vietnam, Nicaragua and Tanzania income poverty statistics

	Poverty line				
	$1 general-CPI	$1 food-CPI	$2 general-CPI	$2 food-CPI	Capability-based
HCR	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93 T, V98 N
IGR	T N V93, V98	T N V93, V98	T N V93 V98	T N V93 V98	V93, T, N V98
PGR	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93 T V98, N
Sen	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93 T V98, N
FGT(1.5)	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93 T V98, N
FGT(2, 2.5, 3)	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93 T N V98
FGT(3.5)	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93, T N V98
FGT(4)	T N V93 V98	T N V93 V98	T N V93 V98	T N V93 V98	V93, T N V98

Notes: N stands for Nicaragua 1998, T stands for Tanzania 2000/1, V93 for Vietnam 1993, and V98 for Vietnam 1998. For FGT(3), under the capability-based poverty line, T is not significantly different from N. Under the capability-based poverty line, FGT(3.5) and FGT(4) of Tanzania can be deemed to be larger than corresponding measures of Nicaragua only at the 10% significance level.

Table 1.9. Synthetic world A (Nicaragua 1998, Tanzania 2000, Vietnam 1998; population 115,027,080)

	Poverty line		
	$1/day	$2/day	Capability-based
World A head count (HC)	31,529,871.55	67,851,421.34	42,252,195.8
World A HC ratio	27%	59%	37%
Nicaragua's share of world A HC	7%	6%	3%
Tanzania's share of world A HC	81%	47%	32%
Vietnam's share of world A HC	13%	47%	65%

a money-metric poverty line. There is a straightforward way to understand this phenomenon. Income poverty estimates are determined by the level of the poverty line and the income profile (or distribution of absolute incomes) in each country. A shift from the $1 per day IPL to the $2 per day IPL entails a doubling of the poverty line in each country (since the PPP used to convert the IPL into local currency and the CPI used to convert the poverty line from the base year to the assessment year do not change as a result of this shift). Although such a shift need not preserve ordinal rankings of poverty across countries (since income profiles can vary in shape across countries, so that the impact of the doubling of the poverty line on the headcount may vary from country to country) it has done so in this case. In contrast, a shift from a money-metric ($1 or $2 per day) IPL to a capability-based poverty line entails a proportionate change in the magnitude of the poverty line which varies in proportion from country to country. For example, a shift from the $1 per day poverty line to the capability-based poverty line leads to an increase in the poverty line by 84 per cent in Vietnam in 1993 whereas it leads to a decrease of 45 per cent in Tanzania in 2000/1. The shift from money-metric to capability-based income poverty lines leads to changes that vary both in direction and magnitude from country to country. It is not surprising that the ordinal rankings of income poverty estimates of countries change as a result. A single correction factor applied to the money-metric poverty line in all countries will not cause the money-metric poverty line to generate similar result as a capability-based concept of income poverty.

The third question we asked was whether the estimated extent of aggregate income poverty and the contribution of a specific country to aggregate income poverty are influenced by the criterion used to identify the poor. Since the poverty estimates vary so much, it is not surprising that both aggregate income poverty and the share of that aggregate represented by income poverty in each country are affected. In Tables 1.9 and 1.10, we generate "synthetic" worlds consisting of just three countries. Synthetic world A consists of Nicaragua in 1998, Tanzania in 2000, and Vietnam in 1998. In Synthetic world B, we have Nicaragua in 1998,

Table 1.10. Synthetic world B (Nicaragua 1998, Tanzania 2000, Vietnam 1993; population 108,855,380)

	Poverty line		
	$1/day	$2/day	Capability-based
World B head count (HC)	36,955,134.83	80,554,709.27	55,901,134.61
World B HC ratio	34%	74%	51%
Nicaragua's share of world B HC	6%	5%	3%
Tanzania's share of world B HC	69%	40%	24%
Vietnam's share of world B HC	25%	56%	73%

Tanzania in 2000, and Vietnam in 1993. The synthetic worlds are based on the actual populations of these countries in these years. Both the extent of aggregate income poverty and the contributions of each country to aggregate income poverty do indeed vary significantly according to the criterion used to identify the poor. In both synthetic worlds, a capability-based analysis leads to a worldwide headcount ratio which is substantially different from those generated by the $1/day and the $2/day identification criteria, and is between them. The contribution of individual countries to global income poverty varies dramatically depending on the identification criterion used. For example, in the first artificial aggregate considered, Vietnam's share of world income poverty rises from 13 per cent (using the $1/day identification criterion) to 65 per cent (using the capability-based identification criterion).

Our rankings of countries must not in any way be taken as authoritative. Our results suffer from many obvious flaws, among which are the following. First, the survey designs are different in different countries, forcing us to make certain judgements in order to carry out this exercise, and these judgements may be questioned. Second, the non-food poverty line we construct (based on the equiproportionality assumption) may be inappropriate, and indeed its appropriateness may vary from country to country. Third, we do not use equivalence scales to adjust for differences in the calorie and other requirements of different groups of people (as defined by sex, age, etc.). Fourth, while it is useful to employ the consumption pattern of a reference quintile in order to define the composition of the food basket assumed necessary to command at the poverty line (in order to make appropriate allowance for prevailing food habits and preferences), this procedure may also lead to problems arising from the existence of differences in real income across countries. If the reference quintile in one country possesses a higher real income than that in another, it may also possess a richer diet (e.g. one that is more varied and contains foods that are nutritionally or otherwise superior). This reference quintile may consume more "expensive calories" than does that in another country, and hence the food poverty line imputed by our procedure in this country would be (arguably

inappropriately) higher. The result would be a substantive non-equivalence of the poverty line across countries, which may be thought to undermine the claim that we have established comparable poverty lines.

Concerns of this type are legitimate. However, such problems can be diminished or overcome in a more comprehensive and detailed future program of poverty line construction and survey design aimed at more adequately supporting capability-based income poverty comparisons. We have referred to the procedure used as "capability-based" in only the most limited sense, with the goal of highlighting the conclusion that even such a limited approach can have results which are very different from those of the money-metric approach.

V. Conclusions

A requirement for meaningful comparison and aggregation of poverty across countries is that the same criterion must be used to identify the poor regardless of where they live. We have argued that the use of an identification criterion based on the possession of elementary capabilities provides an approach to international income poverty comparison and aggregation that is both coherent and meaningful, unlike existing money-metric approaches. In our empirical exercise involving three countries from three continents (Nicaragua, Tanzania and Vietnam), we have demonstrated that it is possible to produce internationally comparable capability-based income poverty estimates of a limited kind using existing data sources. Standard errors were constructed and intersection partial ordering techniques were employed to establish which pair-wise inter-country poverty comparisons are robust with regard to the choice of identification criterion and which are not. In our case study, both cardinal and ordinal comparisons were affected by the choice of approach.

This finding suggests that the choice of an identification criterion may be an important determinant of our judgements concerning which countries are poorer than others and by how much. We do not make the claim that our poverty estimates are authoritative, because they were produced using data sources that were not specifically designed to support the exercise we have undertaken and they are based on strong simplifying assumptions. However, unlike existing money-metric international poverty lines, our poverty lines possess a more meaningful and uniform substantive interpretation. The fact that they lead to estimates of absolute and relative poverty levels different from those tied to money-metric poverty lines suggests that existing methods of global poverty estimation ought to be critically re-evaluated.

The exercise presented here points to the desirability of undertaking international coordination of survey design and poverty line construction methods. Such coordination will facilitate wide application of a more comprehensive form of capability-based international income poverty comparison and aggregation than that attempted here. An effort of this kind must identify relevant elementary capabilities and the characteristics of the commodities that promote them. There may be almost universal agreement on some elementary income-dependent capabilities (such as the ability to be adequately nourished) and on the characteristics of commodities that promote them (such as calorie content), whereas agreement about other relevant elementary income-dependent capabilities (and the characteristics of commodities that promote them) may not be so readily achieved. The possibility of controversy over what the relevant elementary income-dependent capabilities are and how they are furthered is not in itself reason to dismiss the approach as infeasible. Rather, it is reason to seek an operationally adequate consensus over such questions.

Although our aim has been to show the feasibility and desirability of undertaking capability-based income poverty comparisons using available data, we have not meant to suggest that available data are adequate for this purpose. The development of common protocols for international survey design and poverty line construction is a requirement for increasing the coherence and meaningfulness of international poverty comparison and aggregation. Finally, income poverty assessment, although an essential dimension of poverty assessment generally, remains only one aspect of such assessment, as rightly noted by the advocates of the capability perspective. Income poverty assessment must be informed by that perspective but is a limited domain of application which cannot begin to exhaust the relevance and reach of that transformative approach.

References

Altimir, O. (1982), *The Extent of Poverty in Latin America*, World Bank Staff Working Papers (Washington, DC: World Bank).

Andrews, D., and Buchinsky, M. (2000), "A Three-Step Method for Choosing the Number of Bootstrap Repetitions", *Econometrica*, 68(1): 23–51.

Deaton, A. (1997), *The Analysis of Household Surveys* (Baltimore: Johns Hopkins University Press).

Howes, S., and Lanjouw, J. O. (1998), "Does Sample Design Matter for Poverty Comparisons?", *Review of Income and Wealth*, 44(1): 99–109.

Lancaster, K. J. (1971), *Consumer Demand: A New Approach* (New York: Columbia University Press).

National Bureau of Statistics, Tanzania (2002), *Household Budget Survey 2000/2001* (Dar es Salaam: National Bureau of Statistics, Tanzania).

Reddy, S., and Pogge, T. (forthcoming), "How Not to Count the Poor", in J. Stiglitz, S. Anand and P. Segal (eds), *Debates on the Measurement of Poverty* (Oxford: Oxford University Press).

Sen, A. (1981), *Poverty and Famines* (Oxford: Clarendon Press).

—— (1999), *Development as Freedom* (New York: Anchor Books).

World Bank (1999), *Vietnam Development Report 2000: Attacking Poverty*, World Bank Report 19914-VN (Washington, DC: World Bank).

—— (2001), *Nicaragua Poverty Assessment: Challenges and Opportunities for Poverty Reduction*, vol. 1, World Bank Main Report 20488-NI (Washington, DC: World Bank).

This page intentionally left blank

CHAPTER 2

THE CAPABILITY APPROACH AND THE POLITICAL ECONOMY OF HUMAN DEVELOPMENT

AMIYA KUMAR BAGCHI

I. Introduction

Amartya Sen has pioneered the patient approach to human development of continually searching for ways in which the capabilities of *homo sapiens* can be extended to their fullest potential. He has also used this approach to provide detailed analysis of capability deprivation, especially in the poorer countries of the world. In a related effort, he has expanded the meaning of rationality beyond the typical Chicago School notion that it is self-interest, defined quite narrowly, that drives human beings and provides the structural and institutional features of society (Sen 1977, 1993a). While he focuses on individuals as the agents of social activity, the individuality he works with is social individuality (Bagchi 1998a). In yet another strand of his work, along with such philosophers as Hilary Putnam, he has broken

down the positivist binaries of fact versus value, and the conventional academic binaries of theory versus practice (Sen 1982: part 3; Putnam 2002). The triptych of books he edited jointly with Jean Drèze have the title *The Political Economy of Hunger* (Drèze and Sen 1990–1). I want to explore further the intimate connections between the capability approach and the political economy approach, which was the staple of classical economists from William Petty to Adam Smith, John Stuart Mill and Karl Marx before the rather peculiar separation of economics from politics and philosophy took hold in mainstream economics. I will use the term "critical political economy" to signal the point that the genre of work I am talking about brings a critical attitude to all human institutions, such as the state, the nation, the family, the community and the international political order. The two greatest practitioners of critical political economy, Marx and Engels, regarded the earlier corpus of political economy as an analysis of the capitalist system, and Marx regarded most of his own work as a critique of that corpus. I am following that lead in using the epithet "critical political economy" (Marx 1859/1970, 1867/1887). Critical political economy addresses questions which do not presuppose that capitalism is the only possible human destiny.

The capability approach and critical political economy share a common genealogy from Aristotle to Adam Smith, Mary Wollstonecraft, John Stuart Mill, and Marx and Engels. The purpose of the present chapter is not to trace that genealogy but to show the ways in which a combination of the two approaches can enrich the analysis of some of the most pressing issues facing students of the human sciences.

II. The Analysis of the Complexity of Human Capabilities and the Dialectics of Changes in the Corpus of Critical Political Economy

In their "Introduction" to *The Political Economy of Hunger*, Drèze and Sen write:

> the dominant tradition of economics is much narrower now than it was in the classical political economy of Adam Smith, Robert Malthus, David Ricardo, Karl Marx, John Stuart Mill, and others. Thus the old and archaic term for economics as such is also a reminder of the breadth of the earlier tradition of the subject. Many of the analyses of the kind that are now seen as interdisciplinary would have appeared to Smith or Mill or Marx as belonging solidly to the discipline of political economy as a subject.

It does not really matter whether political, social, and cultural influences on economic matters are counted inside or outside of the discipline of economics, but it can be tremendously important not to lose sight of these influences in analyzing many profoundly important economic problems. (Drèze and Sen 1990: I, 2–3)

Sen has done much to refresh the analytical and conceptual apparatus required for constructing a full-blown, critical political economy. On the other side, the critical political economy approach can bring into focus certain issues that have not been a major concern of social scientists who follow the lead of Sen, although Sen's own work has addressed some of those concerns. One of the best expositions of the critical political economy approach was given in Marx (1859/1970: 205–7):

When examining a given country from the standpoint of political economy, we begin with its population, the division of the population into classes, town and country, the sea, the different branches of production, export and import, annual production and consumption, prices, etc.

It would seem to be the proper thing to start with the real and concrete elements, with the actual preconditions, e.g., to start in the sphere of economy with population, which forms the basis and the subject of the whole social process of production. Closer consideration shows, however, that this is wrong. Population is an abstraction if, for instance, one disregards the classes of which it is composed. These classes in turn remain empty terms if one does not know the factors on which they depend, e.g., wage-labor, capital, and so on. These presuppose exchange, division of labor, prices, etc. For example, capital is nothing without wage-labor, without value, money, price, etc. The seventeenth-century economists, for example, always took as their starting point the living organism, the population, the nation, the State, several States, etc., but analysis led them always in the end to the discovery of a few decisive abstract, general relations, such as division of labor, money, and value. When these separate factors were more or less clearly deduced and established, economic systems were evolved which from simple concepts, such as labor, division of labor, demand, exchange-value, advanced to categories like State, international exchange and world market. The latter is obviously the correct scientific method. The concrete concept is concrete because it is a synthesis of many definitions, thus representing the unity of diverse aspects.

The primitive analytical elements of critical political economy are real human beings with their lived experience and not some construct called Western man or Eastern man (Marx and Engels 1845–6/1976, esp. 41–3), classes determined by the relative degrees of control over the means of production, states that organize the means of coercion and security needed to overcome irreconcilable private interests, communities to which particular groups of people owe their allegiance or are supposed by others to owe such allegiance, the nation as the overarching suzerain over communities providing legitimacy to the state, relations of dominance and subordination in the international economic and political order—relations that are often called colonialism and imperialism. Critical political economy recognizes that class and community structures can change with time, as do the nature of the other elements that go into its making (for expositions of some of these elements, see Sweezy 1946; Ball and Farr 1984; Foley 1986; Wright 1997). But the process of

change in the long run is a dialectical one, in the sense that the structural relations in one epoch or one segment of reality are themselves changed by seismic changes in internal or international economic and political factors, so that no linear dynamics can be traced out by simply examining some "purely economic" equations. But in an ideal-typical capitalist system, there are some regularities that keep cropping up during the period in which that system has come into dominance. I will sketch only a few of these regularities.

Sen's method has throughout been to recognize the complexity of human relations and the complexity of the notion of rationality that must inform judgements about human welfare and policy analysis. As I have pointed out (Bagchi 2000), the young Marx and Engels had also a conception of the potential of human beings and their thrust towards the realization of that potential, which anticipates many of Sen's concerns about freedom as the ultimate frontier of human capabilities.

Marx and Engels (1845–6) regarded the self-activity of labor as one of the defining characteristics of human beings who have overcome alienation and have become free (for an exposition of the Marxian notion of alienation in all its dimensions, see Ollman 1971). The human struggle for freedom is also a struggle to master nature while recognizing that human beings are themselves part of nature. Freedom from worries about animal needs such as adequate food and shelter requires conscious labor and is one of the major propulsive forces of history (Cohen 1988). The human capabilities whose nature Sen has done so much to illuminate are expanded by the growth of productive forces. But that growth is conditioned by the social relations in which human beings are involved once they have acquired tools and culture that distinguish them from the great apes. One of the chief concerns of critical political economy is to illuminate the ways in which productive forces shape human relations and are in turn shaped by them. Whether this relation is harmonious or conflictual depends on the historical context.

Marx and Engels were highly critical of the capitalist system even though they regarded it as an advance over all preceding socio-political systems. That criticism was based on several grounds. First, capitalism impedes the self-development of human beings. Second, it impedes the development of productive forces and leads to such a distribution of the social product as to throttle the human potential of the true workers. All inegalitarian social systems require a coercive state apparatus. But since capitalism is based on atomistic competition driven by the profit motive, it also pits state against state and provokes wars as a means of competition.

Sen's "The Profit Motive" (Sen 1983) is devoted to the examination of the ethics and consequences of the uninhibited pursuit of the profit motive. Adam Smith's example of the baker, the butcher, etc. providing us with our daily bread and meat not because of their benevolence but because of their self-interest is quoted *ad nauseam* by devotees of the unregulated market to prove that markets driven by the profit motive always produce the best of all possible worlds. Sen challenges that view, partly because it is a selective view of the way that individual actions

collectively produce results not intended by any particular agent. Market exchanges can also produce undesirable results that no particular agent could have foreseen. Smith's "invisible hand" (a widely misunderstood metaphor, as shown by Muller 1993: ch. 9; Rothschild 2001: ch. 5) is only one of many examples of unintended consequences of individual or collective action. The paradox of thrift, analyzed by Keynes (1936), is another example of the folly of everybody trying to save when unemployment and underutilization of capacity are rife.

The so-called second fundamental theorem of neo-classical economics, namely, that "every Pareto optimal outcome can be reached through some competitive market equilibrium corresponding to some initial distribution of 'endowments' or resources owned" (Sen 1983: 94), is based on the assumption, *inter alia*, that the production or exchange process must be characterized by constant returns to scale and the absence of any externalities. But the endowments may be such that the Pareto-efficient equilibrium leaves some persons wallowing in obscene luxury while others are starving.[1] Sen (1983: 94) points out that an ethically attractive competitive equilibrium may "involve a political process—indeed, quite possibly—a totally revolutionary one requiring a thorough redistribution of the ownership of means of production".

Sen goes on to argue that competitive markets cannot support the exchange of "positional goods", that is, goods whose enjoyment depends "on the relative position of a person *vis-à-vis* others, e.g. a person holding an eminent position in a job hierarchy, or—to take a different type of example—having access to an uncrowded beach" (Sen 1983: 96). To take a contemporary example, Bill Gates, one of the richest men on earth, and George W. Bush, the most powerful political figure in the world, cannot simply exchange their positions by deals in the market. Sen also criticizes the views of free-market advocates such as Robert Nozick (1974), who base their advocacy on the prior right to private property regardless of the disasters that may result from the pursuit of the profit motive based on unregulated private property.

Critical political economy takes on board all of Sen's criticisms of the profit motive as the guiding principle of human activity. It would add basic analyses that Marx and Engels have made of the working of an ideal-typical capitalist system. In their definition, capitalism separates persons into two radically different groups, the first being the owners of means of production, the second being hired by the first to work for a wage in order to produce marketable commodities for the first group to make a profit from. Capitalism is not just any system in which goods are exchanged for money.

Once the system has come to dominate economic and social life, capitalists have little alternative to pursuing profit as their principal goal, as otherwise they

[1] The Arrow–Debreu model of general equilibrium also disallows any uncertainty that is not fully insurable and cannot handle asymmetric information.

might be defeated in the jungle of the market place and sink into the condition of the propertyless proletariat. On the other hand, since the capitalists control all the means of production, workers have to work for those capitalists and have to compete with one another in order to survive (Marx and Engels 1845–6/1976: 34–92; Marx 1867/1887: parts 2–6).

Marx also embedded economies of scale in production, business organization, research, marketing and finance into his account of the way in which capitalism develops. His concepts of "concentration of capital", when a firm grows by expanding its own activities and thereby becomes economically stronger than other firms, and the "centralization of capital", when a firm grows by taking over the businesses of other firms, anticipated by almost a century the mainstream economists' awareness of, and the empirical evidence for, the fact that competition among the few rather than the continuum of agents underlying the pure theory of Walrasian competitive equilibrium is the norm of capitalism and provides the major dynamic forces propelling it. Schumpeter (1950), Steindl (1952/1976) and Baran and Sweezy (1966) took this line of analysis further. In the context of the steeply rising inequality of income and wealth from the 1980s, this genre of work has acquired added significance (Bloch 2000).

More generally, many publicists, novelists, poets and doctors in the nineteenth century were aware that while the Industrial Revolution was increasing the productivity and average incomes of people in the leading industrializing countries, such as England, Belgium, Germany or the USA, not everybody's income was increasing and more importantly, the quality of life of even those whose incomes were increasing was deteriorating in important respects. Marx and Engels were among the leading analysts of the contradictory directions in which human development was pulled by the Industrial Revolution (see e.g. Engels 1845/1969). In fact, in England, the USA and Canada, the longevity and/or heights of men and women declined in some of the peak decades of industrial advance and in many cases literacy, especially that of women, declined in those years (Bagchi 2005: chs 7 and 8). Moreover, in the areas that were fully or partially colonized by the European countries, the nineteenth century witnessed huge famines, which were most virulent, ironically enough, in the decades in which Western Europe and the overseas offshoots of Europe were undergoing an epidemiological transition (Bagchi 2005: ch. 18).

The evolution of modern capitalism has also been associated with massive market failures—markets that do not exist or are characterized by private provisioning at a cost that the poor cannot afford. Markets that have been called "obnoxious" (Kanbur 2001) such as those in opium, tobacco and tobacco products, and arms, grew under the patronage of states strongly supportive of profit-driven merchants and producers.

Market failures and the virtual absence of public provisioning for public goods in colonial countries left the subject populations poor, illiterate and

short-lived. Although the situation improved after these countries were liberated, social constraints and niggardliness of public expenditure soon slowed down or even reversed the gains in education and health. The work of Sen and his collaborators has focused on both these areas of capability deprivation. Students of critical political economy would claim that these cases of obnoxious, malfunctioning or absent markets were not mere happenstance but were systematic outcomes of the way states influenced by profit-making merchants or by the elite in ex-colonial countries shaped public policies. Britain, the most powerful nation in the world at the time, forced a war on China in 1841 because the latter refused to legalize the import of opium, in which British merchants were the principal agents. Arms were used by slave traders in Africa both to capture Africans to be shipped across the Atlantic and as an item of merchandise coveted by the African intermediaries in the trade. Arms still constitute one of the major items of merchandise, with the United States the top trading nation in that sector.

I shall very briefly sketch the way oligopolistic markets and failures of public provisioning have gravely endangered the state of health of some of the poorest people in the world. The epidemiological transition in the affluent countries of today took place roughly between the 1870s and the 1930s. The major facilitators of this transition were better public sanitation, the provision of clean water, a cleaner urban environment, higher levels of nutrition, and preventative measures against killer diseases such as smallpox, typhoid, cholera and plague (Bagchi 2005: ch. 7). Medicines, especially those produced for profit, played little role in this. The situation began to change with the discovery of sulpha drugs in the interwar period. But the real growth of private medical companies, led by such giants as Pfizer, Hoffman-La Roche and GlaxoSmithKline, occurred after World War II. The rise of antibiotics from the 1940s and biotechnology products from the 1980s on the one hand, and the so-called lifestyle diseases of the better-off sections of population on the other, made drugs and pharmaceuticals companies some of the most profitable enterprises in the world.

The delivery of health care involves asymmetric information, moral hazard, adverse selection and opportunistic behavior to a much greater extent than almost any other service, a fact that has been brought out by economists such as Kenneth Arrow, Joseph Stiglitz and others. Almost all Western European countries and Canada and New Zealand, not to speak of countries of the erstwhile Soviet bloc, had instituted a health service of one kind or another. Among the affluent countries, the USA is practically alone in leaving most health care and its financing in the hands of private corporations. But governing the market in health care, so as to ensure health for everybody, has proved to be very tricky everywhere and impossible in poor countries (Bagchi 2007b). The problems stretch from the lack of incentive for big drug companies to spend money on R&D for treating the diseases of the poor, and even more to find vaccines or preventatives for various new diseases such as HIV/AIDS or new forms of old diseases, such as drug-resistant malaria, to

their excessive allocation of funds for advertising compared with R&D and careful clinical testing (Hollis 2007; Sloan and Hsieh 2007). And these problems are compounded by the systematically opportunistic behavior of drug companies, charging excessive prices to the unwary or unprotected, trying to suppress unfavorable results of drug trials, pretending that slight variations in formulations without additional therapeutic value are an improvement that should be protected by patents and for which much higher prices should be charged, and corrupting doctors with various benefits, including outright bribes, to prescribe useless or positively harmful drugs at exorbitant cost (Angell 2004).

The current patent regime, encoded in the TRIPS clauses of the WTO agreement, and proving to be a deterrent both to innovation and diffusion of new drugs (Chaudhuri 2005), was the result of many years of systematic effort and lobbying by big corporations led by Pfizer, the biggest drug company in the world (Drahos and Braithwaite 2003).

While the private health care sector has proved to be totally incapable of ensuring health for all in poor countries, the state has been badly weakened financially, and in other ways, in these countries, especially after the onset of financial liberalization, debt crises and structural adjustment measures in Latin America, the Caribbean, most countries of Africa, and South and South-East Asia. On top of that, many of these countries have suffered massive capital flight and brain drain, including the export of doctors trained at high cost in such countries as Ghana and Malawi (GCIM 2005).

The capability approach has also enriched our appreciation of the complex dimensions of health care, nutrition and survival in a healthy state. Sen's emphasis on different requirements for differently placed individuals—for which he has acknowledged Marx's insistence, in his critique of the Gotha Program of the nascent Social Democratic Party of Germany, on different provisions to be made for persons with different needs (Marx 1875/1970)—has enriched our understanding of why the needs of women, children, and the mentally and physically challenged have to be examined with special care. A very large proportion of women in poor countries live under some kind of patriarchy and are deprived of property rights and access to education, nutrition, health care and employment with dignity. Exclusive emphasis on increasing income even in countries with apparently socialist ideologies or relatively egalitarian social structures, such as China and South Korea, has resulted in high degrees of discrimination against women in work opportunities and earnings, and discrimination extends even to the equal right of survival of girls and boys. The capability approach can bring such dark realities of poverty into the light of day.

The critical political economy approach can exploit another aspect of the changeability of the human condition emphasized by Marx. In a speech by Marx to the First International Working Men's Association in June 1865, Marx (1865: ch. 14) stated:

The value of the laboring power is formed by two elements—the one merely physical, the other historical or social. . . . A quick succession of unhealthy and short-lived generations will keep the labor market as well supplied as a series of vigorous and long-lived generations. Besides this mere physical element, the value of labor is in every country determined by a *traditional standard of life*. It is not mere physical life, but it is the satisfaction of certain wants springing from the social conditions in which people are placed and reared up. The English standard of life may be reduced to the Irish standard; the standard of life of a German peasant to that of a Livonian peasant.

Thus the kind of concern Sen (1999: 73–4) attributes to Smith about defining "necessities" acquires a social and historical dimension in Marx's work (see also Miller 1991).

III. Capability, Identity and the Dark Side of Globalization

Sen has used his critique of the theory of human action based solely on self-interest and self-welfare to create a space for commitment, that is, action that is not driven by "self-centered welfare", the "self-welfare goal" or "self-goal choice" (Sen 1977, 1985/2002, 2005: 6). That commitment can be to an abstract sense of justice, to a program for political or social change, to a business group and its ethos, to a religious or social community. But the commitment can be directed towards the forcible denial of freedom of choice to others. It is against that confining of people's choice and the denial of the fact that we all have plural identities that a large part of Sen's recent lectures and writings have been directed (see e.g. Sen 1999b, 2006).

People have multiple, and changeable, identities. Capitalist competition itself creates new identities, as it demotes some sections of workers from skilled to unskilled status, and changes the rankings of different groups of workers (Botwinick 1993). The overall increase in the concentration of industry and structural change in the composition of industrial output have led to increased wage inequality among workers in many countries (Galbraith and Berner 2001).

The support of plural identities requires the substantive freedom whose elaboration is as integral a part of the capability approach as its insistence on procedural freedom. In Sen's illustration of multiple affiliations, we can picture a woman as a vegetarian, a lawyer, a lover of jazz, a heterosexual, and a supporter of gay and lesbian rights (Sen 2006: 45–6). But a common, often homeless, worker in an Indian town has no option of listening to Tagore songs or reading the fiction of Bibhutibhusan Bandyopadhyay. If that worker happens to be a Muslim, he might go to a mosque, and get some food and even shelter for the night. It is not enough

to have a radically individuated society with no connection between human beings except that of accidental affiliation, a society in which the demand for a private domain also includes the recognition of private property in any number of assets, commodities and capabilities. That society, as Marx argued in his article on the Jewish question (Marx 1844), will pit one individual against another in ceaseless competition. Competitors will use attributed identities to enslave workers or to aggrandize themselves. That factor has been one of the principal reasons why, contrary to the dream of liberal philosophers working out the political and moral foundations of a capitalist society along the lines of freedom of individuals, the world remains mired in identitarian politics.

Sen has discussed the ethics of market exchange and finance in many places (see e.g. Sen 1993b, 1993c, 1999a: ch. 5). From the point of view of either the critical political economy approach or the capability approach, these are human institutions and the consequences of their operation have to be judged in accordance with whether they advance or pull back the frontier of human potential. Since these institutions have various effects on different groups of people, distributional judgements are involved in assessing their consequences. Financial innovations of the type that go by the name of financial liberalization, which have been adopted by the leading economic powers and have often been forced upon the poor, debt-ridden countries of Africa, Latin America and Asia, have enriched the wealthy to previously unimaginable levels. In almost every country, financial liberalization has also led to enormous increases in inequality. It has also been associated with severe declines in growth rates of per capita incomes and has often led to absolute declines in income per head, the decimation of social security systems and declines in real expenditure per capita on education and health because of the withdrawal of the state from most social and productive sectors (Bagchi 2005: chs 22–4; Bagchi 2007a). For the first time since the beginning of the epidemiological transition in Western Europe, life expectancy in a swathe of countries in Sub-Saharan Africa and several countries in the former Soviet Union has declined, in some cases to 35 years or so, the same as at the dawn of their liberation from colonial rule. While the capability approach should be used to grasp the multidimensional nature of this human tragedy, critical political economy can be used to see why, despite the arguments of many economists (e.g. Stiglitz 2002, 2006), the pleadings of many UN bodies, and world-wide protests against what goes by the name of (G7-led, corporatist) globalization, the order prescribed by the IMF, the WTO and the World Bank continues to govern the lives of the world's billions.

While the complexity of human relations and individual human capabilities will continue to be mapped in their many dimensions, the substantive freedoms that undergird the expansion of human capability must be continually tracked and not ignored in the name of more sophisticated methods. For example, there is a good deal of debate, taking its point of departure from the work of Pitambar Pant, Anthony Atkinson, Amartya Sen, James Foster, Angus Deaton, Abhijit Sen,

Himangshu and others (for references, see Himangshu 2007), about how exactly to define and measure the degree of poverty in India. But the officially measured number of people below the poverty line in rural areas in India over the period of the neo-liberal economic reforms looks very peculiar since, over the same period, the availability of food grains and of calories delivered by food of all kinds has also declined, and the calorie intake of people below the official poverty line has fallen further and further below the calorie intake norm defining the poverty line (Patnaik 2007: Chart 4a, Table 2). Even in the state of Kerala, which has been lauded as a success in human development in a poor region, the calorie intake has gone down over this period.

From 1975, in collaboration with UNICEF, the Indian government has been implementing an innovative program called the Integrated Child Development Scheme (ICDS). Its geographical coverage has been extended over the years and by 2006 it had come to cover notionally about one quarter of the Indian population. Despite the existence of such a scheme, levels of child malnutrition remain very high, higher than in many countries of Sub-Saharan Africa (Shiva Kumar 2007). A. K. Shiva Kumar, Advisor to UNICEF in India, makes use of the data collected by the third round of the (Indian) National Family Health Survey (NFHS3), carried out in 2005/6 to measure the access of children, under the ICDS and other schemes, to supplementary nutrition, to immunization, and so on, and finds that their access is very limited in most parts of rural India. He also tabulates the extent of undernutrition, as measured separately by children who are underweight, children who are wasted, and children who are stunted. According to those data, 49 per cent of children are underweight, with lower proportions of children being wasted or stunted. The situation may be even grimmer than that. Svedberg (2002) defines a comprehensive index of anthropometric failure (CIAF) which measures the number of children who suffer from at least one of the three kinds of anthropometric failure. On the basis of data from NFHS2 (the second round of the National Family Health Survey), carried out in 1998/9, Nandy *et al.* (2005) find that just about 60 per cent of Indian children below age 3 suffered a CIAF condition. They also find a strongly negative relationship between the proportion of children suffering from CIAF and the standard of living of the households covered.

The political economy approach helps us understand the very partial success of the ICDS program against the background of economic reforms that badly damaged the public distribution system of food and public expenditure on agricultural investment, and squeezed the employment conditions of the rural population. The capability approach can then examine the system of allocation of nutrients within a typical poor Indian family, in which a woman has restricted access to gainful employment and has less autonomy in deciding how to spend the money she earns even when she is an earning member of the family. Will even a pregnant mother agree to take supplementary vitamins and food only for herself if her children above

the age of 3 or 6 (depending on the cut-off point for public provision) are going hungry?

In 2000, at the dawn of the new millennium, the United Nations proclaimed a number of Millennium Development Goals (MDGs), to be attained by 2015. They include significant reduction in extreme poverty, as measured by those living below an income or exchange entitlement of one dollar a day, the universalization of primary education, preferably by 2015, a significant reduction in gender disparities in education and earnings, progress towards sustainable development, and a global partnership in development. It was obvious by 2006 that most of these goals were not going to be reached by that date. The sexual and reproductive health of women is extremely important for attaining these goals: they require the empowerment of women to take control of their own lives. Under the political and social order prevailing in poor countries, their states or communities or private business sectors are not able to provide the necessary resources, incentives and social environment. Under the international economic and political order, especially after the emergence of the WTO, which seriously discriminates against the agriculture of the poor countries, the progress of empowerment is impossible. The UN MDG report on sexual and reproductive health has confirmed these apprehensions (UN 2006). The major report in 2007 on progress towards attaining the MDGs does not sound more hopeful (UN 2007). While the MDGs include an enormous growth in aid from rich countries to poor countries, actual aid has declined. Deforestation in poor counties continues apace. Critical political economy analysis could have predicted all these disappointing outcomes. Why should the rich countries, whose international economic policies have been increasingly guided by the interests of their own business corporations, pour aid into countries whose per capita incomes are too low to repay the aid by creating a big market? How can poor countries stop deforestation, when very often the survival of the poor requires them to scrounge from the soil, and when under various legal agreements, business strategies have been freed of any social responsibility, except in a few Western European economies? Practitioners of critical political economy have suggested that the poor in all countries should struggle for changes in policy orientation at national and international levels (see e.g. Foster 2002; McNally 2002; Bagchi 2005). The goals of those struggles can be enriched by drawing on the work emanating from the capability approach (Nussbaum 2000; Agarwal, Humphreys and Robeyns 2004).

Both critical political economy and the capability approach can be used to combat some of the almost deliberately misleading propaganda spread by the organs of the IMF and the World Bank, and by economists unquestioningly accepting their perspective. The portrayal of the "East Asian Miracle" and its downfall are a prime example of the need for such scrutiny. When the World Bank (1993) celebrated the East and South-east Asian miracle, it failed to distinguish the fragility of industrial growth in a country such as Indonesia from its robustness in South Korea.

In fact, the World Bank advised countries like India to emulate the policy models of Indonesia and Thailand rather than that of South Korea (Bagchi 1998b). Moreover, the massive and continuing violations of human rights in Indonesia under Suharto's rule, lasting for almost thirty years, were never mentioned in the World Bank reports. When the Asian economic crisis ensued, the blame was placed on the regimes' lack of transparency and "cronyism", forgetting that if those charges were true, these faults were also responsible for the earlier successes. It turned out that the chief culprits in bringing about the crisis were capital account convertibility and badly regulated financial liberalization. Moreover, the IMF prescriptions as conditions for bail-out loans deepened the depression in the crisis-affected economies (Stiglitz 2002, 2006). Again, a capability approach that pays attention to the effects of a given policy or a crisis can shed better light on the human predicament than can sticking to the aggregative measures of per capita income growth favored by most international organizations. For example, the analysis by Jan Breman (2000) of what happened in rural Java in the wake of the crisis—how employment and incomes collapsed and why the traditional safety nets could not bear the weight—can help us better understand the subsequent political turmoil and inter-religious conflicts in Indonesia than can aggregative figures of alleged recovery (which, by all accounts, has been very fragile).

The capability approach and class analysis are natural complements of each other. Sen (1981: app. B) gives an exemplary demonstration of how class analysis can be used to show the effects of a price rise of food grains on different classes of society and how entitlement failures can occur differently for different sections of society. He uses a two-class model which divides the population between the rich and the poor, the latter being the workers, to bring out the reasons for Malthus's objection to parish allowances being given to workers to protect their real incomes during Britain's war against Napoleon-ruled France. Sen employs a five-class model, distinguishing between (1) agricultural capitalists and landlords, (2) peasants, (3) urban and semi-urban workers, (4) agricultural laborers and (5) rural service providers and craftsmen to bring out the different effects of the Bengal famine of 1943 on different classes. The same framework, with changes in class divisions where needed, can be used to illuminate what happened during the series of famines India had to suffer under British rule throughout the nineteenth century and beyond (Maharatna 1996; Davis 2001). Such an analysis can be embedded in the wider perspective of critical political economy. Why were the British rulers unwilling to spend money to save lives? What were their imperatives during World War II, when the Bengal famine occurred? How did Lord Lytton, then Viceroy of India, set about allocating funds between preparation for the Second Anglo–Afghan War, the ceremonial crowning of Queen Victoria (in absentia) in the Delhi Durbar of 1876, and the money needed to save Indian lives? What were the fiscal imperatives of an empire that maintained a large establishment in London and funded the cost of the whole British army east of Suez?

Sen's work can enrich critical political economy by making it easier to incorporate the ways individuals and groups can differ in their needs, their goals and their commitment. In its turn, critical political economy can help trace out the trajectory of changes in these diversities of experience, commitment, pursuit of goals and actual choice. Such changes, and the political struggles associated with them, can both facilitate and obstruct the processes of democracy in society and polity and thereby stand in the way of advances in human development and the frontier of human capabilities, as the experience of formal democracy, stretching from the USA to India, demonstrates (Assayag 1995; Bagchi 2004; Quadir and Lele 2004; Bagchi 2005: chs 23–4).

References

Agarwal, B., Humphreys, J., and Robeyns, I. (eds) (2004), *Capabilities, Freedom and Equality: Amartya Sen's Work from a Gender Perspective* (New Delhi: Oxford University Press).

Angell, M. (2004), *The Truth about the Drug Companies: How They Deceive Us and What To Do About It* (New York: Random House).

Assayag, J. (1995), *The Making of Democratic Inequality: Caste, Class, Lobbies and Politics in Contemporary India (1880–1995)* (Pondicherry, India: Institut Français de Pondichery).

Bagchi, A. K. (1998a), "Amartya Kumar Sen and the Human Science of Development", *Economic and Political Weekly*, 33: 3139–50.

—— (1998b), "Growth Miracle and its Unraveling in East and South-East Asia", *Economic and Political Weekly*, 33: 1025–42.

—— (2000), "Freedom as Development or as End of Alienation?", *Economic and Political Weekly*, 35: 4408–20.

—— (2005), *Perilous Passage: Mankind and the Global Ascendancy of Capital* (Lanham, Md: Rowman and Littlefield).

—— (2007a), "Finance: Lessons for India", in A. K. Bagchi and G. A. Dymski (eds), *Capture and Exclude: Developing Economies and the Poor in Global Finance* (New Delhi: Tulika).

—— (2007b), "Governing the Market in Health Care: The Social and Political Requirements", in D. McIntyre and G. Mooney (eds), *The Economics of Health Equity* (Cambridge: Cambridge University Press).

Ball, T., and Farr, J. (eds) (1984), *After Marx* (Cambridge: Cambridge University Press).

Baran, P. A., and Sweezy, P. M. (1966), *Monopoly Capital* (New York: Monthly Review Press).

Bloch, H. (2000), "Schumpeter and Steindl on the Dynamics of Competition", *Journal of Evolutionary Economics*, 10: 343–53.

Botwinick, H. (1993), *Persistent Inequalities: Wage Disparity under Capitalist Competition* (Princeton: Princeton University Press).

Breman, J. (2000), *The Impact of the Asian Economic Crisis on Work and Welfare in Village Java*, Die Natalis Address delivered on 12 October 2000 (The Hague: Institute of Social Studies).

Cohen, G. A. (1988), "Restricted and Inclusive Historical Materialism", in G. A. Cohen, *History, Labor, and Freedom: Themes from Marx* (Oxford: Clarendon Press).

Davis, M. (2001), *Late Victorian Holocausts: El niño Famines and the Making of the Modern World* (London: Verso).

Drahos, P., and Braithwaite, J. (2003), *Information Feudalism: Who Owns the Knowledge Economy?* (Oxford: Oxford University Press).

Drèze, J., and Sen, A. (1990–1), *The Political Economy of Hunger*, 3 vols (Oxford: Clarendon Press)

Engels, F. (1845/1969), *The Condition of the Working Class in England*, in K. Marx and F. Engels, *Selected Works*, vol. 3 (Moscow: Progress Publishers).

Foley, D. K. (1986), *Understanding Capital* (Cambridge, Mass.: Harvard University Press).

Foster, J. B. (2002), *Ecology against Capitalism* (New York: Monthly Review Press).

Galbraith, J. K., and Berner, M. (eds) (2001). *Inequality and Industrial Change: A Global View* (Cambridge: Cambridge University Press).

GCIM (2005), *Migration in an Interconnected World: New Directions for Action. Report of the Global Commission on International Migration*, available at <www.gcim.org>.

Himangshu (2007), Recent Trends in Poverty and Inequality: Some Preliminary Results. *Economic and Political Weekly*, 42: 497–508.

Kanbur, R. (2001), *On Obnoxious Markets,* Working Paper WP 2001-08 (Ithaca, NY: Department of Applied Economics and Management, Cornell University).

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money* (London: Macmillan).

Maharatna, A. (1996), *The Demography of Famines: An Indian Historical Perspective* (Oxford: Oxford University Press).

Marx, K. (1844), "On the Jewish Question", Eng. tr. available at <www.marxists.org/archive/marx/works/1844/jewishquestion/index.htm>.

—— (1859/1970), *A Contribution to the Critique of Political Economy*, tr. S. W. Ryazanskaya, ed. M. Dobb (Moscow: Progress Publishers).

—— (1865), "Value, Price and Profit", Speech to the First International Working Men's Association, June 1865; available at <www.marxists.org/archive/marx/works/1865/value-price-profit>.

—— (1867/1887), *Capital: A Critical Analysis of Capitalist Production*, vol. 1, tr. S. Moore and E. Aveling (Moscow: Foreign Languages Publishing House).

—— (1875/1970), "Critique of the Gotha Program", in K. Marx and F. Engels, *Selected Works*, vol. 3 (Moscow: Progress Publishers).

—— and Engels, F. (1845–6/1976), *The German Ideology* (Moscow: Progress Publishers).

Maurer, S. M. (2007), "When Patents Fail: Finding New Drugs for the Developing World", in Sloan and Hsieh (eds) (2007).

McNally, D. (2002), *Another World is Possible* (Winnipeg: Arbeiter Ring).

Miller, R. W. (1991), "Social and Political Theory: Class, State, Revolution", in T. Carver, (ed.), *The Cambridge Companion to Marx* (Cambridge: Cambridge University Press).

Nandy, S., *et al.* (2005). "Poverty, Child Malnutrition and Morbidity: New Evidence from India", *Bulletin of the World Health Organization*, 83(3): 210–16.

Nozick, R. (1974), *Anarchy, State and Utopia* (New York: Basic Books).

Nussbaum, M. (2000), *Women and Human Development: The Capabilities Approach* (Cambridge: Cambridge University Press).

—— and Sen, A. (eds), (1993), *The Quality of Life* (Oxford: Oxford University Press).

Ollman, B. (1971), *Alienation: Marx's Conception of Man in Capitalist Society* (Cambridge: Cambridge University Press).

Patnaik, U. (2007), "Neoliberalism and Rural Poverty in India", *Economic and Political Weekly*, 42: 3132–50.

Putnam, H. (2002), *The Collapse of the Fact/Value Dichotomy, and Other Essays* (Cambridge, Mass.: Harvard University Press).

Schumpeter, J. A. (1950), *Capitalism, Socialism and Democracy*, 2nd edn (London: Allen and Unwin).

Sen, A. (1967), "Isolation, Assurance and The Social Rate of Discount", *Quarterly Journal of Economics*, 81: 112–24; repr. in Sen (1984).

—— (1973/1997), *On Economic Inequality* (Oxford: Clarendon Press).

—— (1977), "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, 6: 317–44; repr. in Sen (1982), 84–106.

—— (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (New York: Oxford University Press).

—— (1982), *Choice, Welfare and Measurement* (Oxford: Blackwell).

—— (1983), "The Profit Motive", *Lloyds Bank Review*, no. 147 (January 1983), 1–20; repr. in Sen (1984), 90–110.

—— (1984), *Resources, Values and Development* (Oxford: Blackwell).

—— (1985), "Goals, Commitment, and Identity", *Journal of Law, Economics, and Organization*, 1: 341–55; repr. in Sen (2002), 206–24.

—— (1993a), "Capability and Well-Being", in Nussbaum and Sen (1993).

—— (1993b), "Markets and Freedoms: Achievements and Limitations of the Market Mechanism in Promoting Individual Freedoms", *Oxford Economic Papers*, 45, 519–41.

—— (1993c), "Money and Value: On the Ethics and Economics of Finance", *Economics and Philosophy*, 9: 203–27.

—— (1999a), *Development as Freedom* (New York: A. A. Knopf).

—— (1999b). *Reason Before Identity,* Romanes Lectures, Oxford University, 1998 (Oxford: Oxford University Press).

—— (2002), *Rationality and Freedom* (Oxford: Oxford University Press).

—— (2005), "Why Exactly is Commitment Important for Rationality?", *Economics and Philosophy*, 21: 5–14.

—— (2006), *Identity and Violence: The Illusion of Destiny* (London, Penguin Books).

Shiva Kumar, A. K. (2007), "Why Are Levels of Child Malnutrition not Improving?", *Economic and Political Weekly*, 42: 1337–45.

Sloan, F. A. and Eesley, C. E. (2007), "Implementing a Public Subsidy for Vaccines", in Sloan and Hsieh (eds) (2007).

—— and Hsieh, C. R. (2007), "Introduction", in Sloan and Hsieh (eds) (2007). *Pharmaceutical Innovation: Incentives, Competition, and Cost–Benefit Analysis in International Perspective* (Cambridge: Cambridge University Press), 1–21.

—— —— (eds) (2007), *Pharmaceutical Innovation: Incentives, Competition, and Cost–Benefit Analysis in International Perpective* (Cambridge: Cambridge University Press).

Steindl, Josef (1952/1976), *Maturity and Stagnation in American Capitalism* (New York: Monthly Review Press).

Stiglitz, J. E. (2002), "The East Asian Crisis: How IMF Policies Brought the World to the Verge of a Global Meltdown", in J. E. Stiglitz, *Globalization and Its Discontents* (London: Allen Lane).

—— (2006), *Making Globalization Work* (London: Allen Lane).

Svedberg, P. (2002). *Poverty and Undernutrition: Theory, Measurement and Policy* (New Delhi: Oxford University Press).

Sweezy, P. M. (1946), *The Theory of Capitalist Development: Principles of Marxian Political Economy* (New York: Oxford University Press).

UN (2006), "Public Choices, Private Decisions: Sexual and Reproductive Health and the Millennium Development Goals", by S. Bernstein and C. J. Hansen; available at <www.unmillenniumproject.org/documents/MP_Sexual_Health_screen-final.pdf>.

—— (2007), *The Millennium Development Goals Report 2007* (New York: United Nations).

Wright, E. O. (1997), *Class Counts: Comparative Studies in Class Analysis* (Cambridge: Cambridge University Press).

CHAPTER 3

INDIA AND CHINA

"THE ART OF PROLONGING LIFE"

LINCOLN C. CHEN

I. Introduction

SHORTLY after rejoining Harvard University twenty years ago, I accompanied Amartya Sen on a field trip to Dhaka, where Marty and I had spent eight of our 14 years of residential work in Bangladesh and India. When we arrived at a BRAC training center in Manikganj, an hour outside of the capital, Amartya proudly reported that this was his ancestral home district. Much animated discussion took place as to how to return Amartya to his home village given the trails, earthen bunds, and riverbeds in between. The most efficient transport, Amartya riding pillion on a motorbike driven by a field worker, was eventually discarded in favor of a very hardy jeep that somehow was able to navigate successfully the narrow paths, returning the man who is arguably East Bengal's most distinguished intellectual product to his home soil.

That trip marked a "Bengali" bonding between Amartya and me, for our love for East Bengal was shared. My most exciting, and possibly best, research on health, population, nutrition and development was conducted in the 1970s in nearby Matlab *thana*, Comilla District. Amartya has always been generous in citing these

I gratefully acknowledge the research assistance of Dominique Lam, who gathered together Sen's writings on India and China and offered useful suggestions, and the research and editorial work of Victoria Fan.

field studies, especially documentation of discriminatory health care and feeding practices against young girls in Matlab families (Chen, Huq and D'Souza 1981).

For the past two decades, Amartya and I have crossed paths and traveled together many times for many activities in many places. A decade at Harvard was devoted to the Center for Population and Development Studies, where Sudhir Anand and we joined to create one of the most intellectually vibrant centers of interdisciplinary research at the university. I recall a marvelous celebration hosted by Sudhir at the Center after Amartya received his Nobel Prize in Economics. With funds from the award, he established two Pratichi Trusts in India and Bangladesh devoted to advancing basic health and education. Pratichi, with the support of Unicef, launched the "Kolkata Group", a forum for policy exchange and advocacy for children's education and health. These ongoing annual forums bring together Pratichi field researchers with scholars, policy-makers, journalists and philanthropists in India and neighboring Bangladesh.

In these and other intellectual exchanges, Sen has invariably illustrated his points with rich examples from India and China. He has underscored the deep historical connections among these two great civilizations (Sen 2004). Exchange, he postulates, began with trade—mostly following the Silk Road. But over time, exchanges expanded in the transfer of knowledge in science, mathematics, astronomy, health and virtually all aspects of life. Accelerated in part by the spread of Buddhism from India in the first millennium, the exchanges tapered off with the decline of Buddhism in the second millennium. Vividly recounted by Sen are the fascinating observations of India by Chinese travelers and the movement of Indian scientists to China in the fifth through eighth centuries (Sen 2004). The sophistication of these two civilizations is well exemplified by the first printed book in the world: a Chinese translation of an Indian Buddhist treatise, the so-called Diamond Sutra, in AD 868, six hundred years before the Gutenberg Bible in Europe.

For Sen, and for many other Indian intellectuals, China has represented a beacon of light illuminating India's own development experiences, successes and shortcomings, just as India's rich culture has deeply permeated into Chinese tradition. Sen has pointed out that thousands of Sanskrit texts have been translated into Chinese. Sen's own study of Sanskrit during his early schooling in his home in Santiniketan no doubt stimulated his fascination with the interplay among these two great civilizations. Many Chinese visitors, short- and long-term, came to visit Rabindranath Tagore in Santiniketan. Somehow, I always felt that Sen's ease in my presence came from his early acculturation to Chinese people and their civilization.

It is the search for contemporary relevance, however, that characterizes the thrust of Sen's inquiries (Sen 1999a: 22–3, 47, 104). Obviously, the sheer demographic weight of India and China is of global importance. Each has grappled with huge development challenges, some of the most dramatic in the twentieth century—war, revolution, famine and epidemics. While it is impossible to capture all of the richness of Sen's comparative studies on India and China, his works have been elegant, pioneering and multidisciplinary—calling upon history, Sanskrit,

philosophy, economics, demography and the health sciences. They have illuminated many issues, such as famine and democracy, human capabilities, gender equity, and freedom and population policies.

Famine and democracy. Stimulated by his first-hand childhood memories of India's Great Bengal Famine of 1943, Sen has examined the Chinese famine during the Great Leap Forward in 1958–61 (Sen 1981). China's famine and independent India's capacity to control famine illustrate one of Sen's astute observations: that famines do not strike democratic polities because an open media exerts powerful political pressures for governmental intervention.[1]

Human capabilities. Sen has pioneered the theory of human capabilities, with basic education and health as core human capabilities essential for exercising choice in the pursuit of freedom (Drèze and Sen 2002: chs 3, 4, 6). In these two basic dimensions of human capabilities—but not necessarily in all others—China has outperformed India, but the exceptional performance of India's Kerala state compared to all of China has generated many illuminating insights (Sen 1991).

Gender inequity. A powerful explanatory force affecting human capability is deeply rooted in both India and China. A strong preference for males in both societies generates an unbalanced sex ratio and a neglect of girls that aggregate into millions of "missing women" (Sen 1990). Sen has been at the forefront of those emphasizing that investment in the education and health of girls yields wide-ranging benefits in fertility decline, family health and economic opportunities (Agarwal, Humphries and Robeyns 2005).

Freedom and population policies. India and China, the world's two most populous countries, offer fascinating contrasts in population policies (Sen 1994). Sen has contrasted "override" (compulsory) population control policies with more "collaborative" (voluntary) approaches to demographic change. While India has fluctuated between both approaches, Sen questions precisely how much of China's dramatic fertility decline can be attributed to its compulsory one-child population policies, as China had already generated so many of the social conditions that induce fertility decline—e.g. basic education and health care.

II. "The Art of Prolonging Life"

Exchanges between India and China, Sen has often noted, have been particularly interested in health care, treating diseases, and, in the words of a Chinese traveler to India, "the art of prolonging life". This historical base has offered Sen a platform

[1] Many feel that Sen should also have been awarded the Nobel Peace Prize for his work on famine.

Table 3.1. Health and development indicators, India and China

Indicator	India	China
Total population, 2004	1,087,124,000	1,307,989,000
GDP per capita (PPP$), 2004	3,139	5,896
Population living below $1 a day (%), 2004	34.7	16.6
Population living below $2 a day (%), 2004	79.9	46.7
Adult literacy rate (% age 15 and older), 2004	61	90.9
Inequality (Gini index), 2004	32.5	44.7
Population in urban areas (%), 2005	29	40
Life expectancy at birth (years), 2004	63.6	71.9
Infant mortality (per 1,000 live births), 2004	62	26
Maternal mortality ratio, adjusted (per 100,000 live births), 2000	540	56

Sources: WHO (2004); UNDP (2006).

to compare the more recent performances of these two countries (Sen and Drèze 1995). What is their current health situation, and how did they get there?

Table 3.1 summarizes contemporary health and development indicators for India and China (WHO 2004; UNDP 2006). These two countries have nearly comparable populations, but China's GDP in purchasing power parity is nearly twice that of India's. Levels of poverty are correspondingly lower in China, but in contrast, India's Gini coefficient for income inequality is much less severe than China's. Adult literacy in China is superior to India's. On average, China is able to "prolong life" for eight more years than India—71.9 versus 63.6. This gap in life expectancy reflects the corresponding differences in infant, child and maternal mortality.

Figure 3.1 plots the paths of child survival against per capita economic growth for India and China between 1960 and 2003 (UNDP 2004). Both countries achieved "independence" at about the same time, with roughly similar levels of child survival. From 1960 through 2003, India improved in child survival and GDP at a fairly steady, albeit moderate, pace. In contrast, China experienced, with some setbacks, accelerating child survival and slow economic growth in the earlier decades, followed by rapid economic growth but more modest gains in child survival in the later decades. Sen has cited China's slow health gains despite robust recent economic growth as reflecting China's weak policy support for health and other social sectors (Drèze and Sen 2002). By 2003, China had greater 0–5 year child survival than India (96.3 per cent versus 91.3 per cent) and higher GDP per capita ($4,730 versus $2,730 in PPP$).

Institutionally, India and China have developed national health systems that are state-financed and state-executed. But their development paths have deviated (Bhalla 1992; Chen 1987; MacFarquhar 2005; Malenbaum 1982). India's constitution designates health as a state subject, and health spending and implementation varies widely among Indian states (Sen 1996). Kerala and much of southern India are quite advanced, while the so-called "bimaru" (Bihar, Madhya Pradesh, Rajasthan and

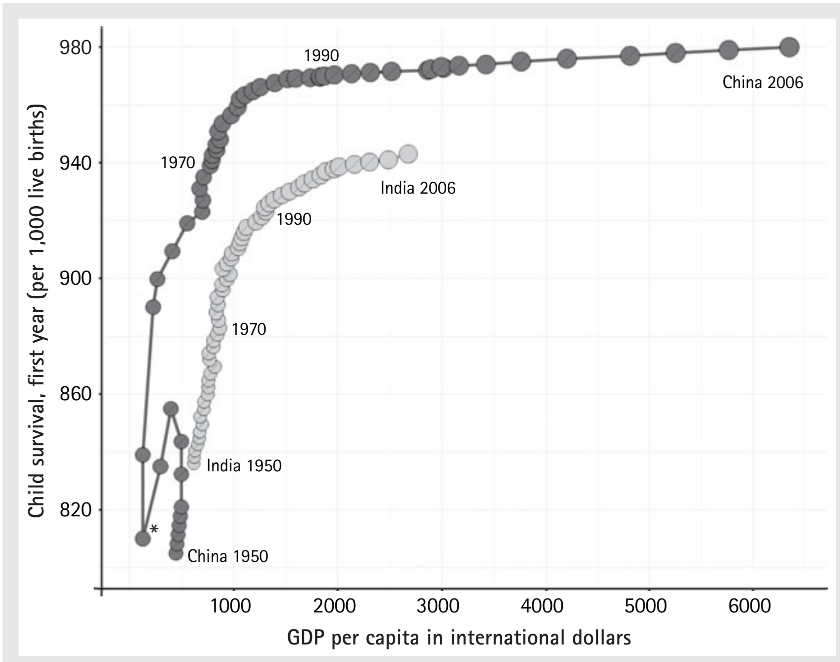


Fig. 3.1. Child survival and GDP per capita, China and India, 1950–1999

Source: <Gapminder.org>.

Uttar Pradesh) are backward. Implementation is dependent on a public bureaucracy, following a colonial tradition. Neither financing nor execution has followed the ambitions at independence for a public-sector health system able to achieve universal coverage of basic services. In policies and regulations, private-sector health activities have been overlooked. China also has a centrally planned health system, with financing and operations delegated to communes and state enterprises. The system attempts to integrate traditional and modern medicine and aims to achieve universal coverage at low cost. Program implementation by a single political party at least initially was more efficiently implemented under an authoritarian state. Evidence suggests that large-scale hygiene campaigns have had some positive effects against infectious and parasitic diseases. Private provisioning and private financing are not permitted by official regulations.

Caution should be exercised in interpreting these data and institutional comparisons. The quality of data, even of those reported by the UN, is not beyond dispute in either India or China. India more than China, however, produces and disseminates many sources of data which are readily accessible to researchers. Whereas triangulation through different data sources can be conducted reasonably easily with Indian data—censuses, surveys, registration, etc.—validating Chinese data is more difficult. Comparisons, especially modest differences, should not be over-interpreted.

III. Looking into the Future

What would a "Sen perspective" on the future of health in India and China look like? Only he, of course, can answer that question. But his intellectual trademark would undoubtedly lead to penetrating insights rooted in history, ranging widely across disciplines, and linked to his concerns over human freedom and human capability.

The "art of prolonging life" in the twenty-first century will not resemble that of the past century—in either India or China. Both have joined the global economy driven by private markets. Unprecedentedly rapid economic growth will increasingly thrust these two countries into the center of the global economy (Huang and Khanna 2003). The role of the private market in the health sector is likely to expand, given the inevitable role of markets in new technologies, the economic demands of health professionals, and commercial opportunities for the pharmaceutical, insurance and hospital industries. Unless public interventions are more effective than they have been, income inequalities are unlikely to improve and health inequities may unfortunately worsen. Issues of access to basic services, protection against financial catastrophes due to illness, and popular satisfaction—especially for subgroups who are poor, rural, minorities, or otherwise disadvantaged—are likely to remain challenges.

Political rhetoric on health, implemented through recent policy reforms, is growing in both India and China. India has launched "rural health missions" that will develop village-based cadres of health workers who bring basic health services to rural populations (Kumar 2007). China is expanding its "rural cooperative health insurance" to achieve universal coverage by 2010. In both countries, even as these rural health efforts are mounted, high-tech tertiary urban hospitals are growing by leaps and bounds, many driven by profit incentives (Blumenthal and Hsiao 2005).

Moving beyond their economically hindered pasts, India's and China's rapidly expanding economies are generating far greater fiscal capacities, public and private, that could be applied to solve these health sector challenges. In such pursuits, many choices will have to be made and many paths will be possible. All, however, will be influenced by three major forces—changing demographics and epidemiology, reforms of the health sector, and the development dimensions of health change.

III.1 Demography and Epidemiology

Given current population growth rates, India will overtake China as the world's most populous country around 2035—each will have around 1.5 billion people, together approaching nearly 30 per cent of the world's population (UN 2005). Both countries are rapidly urbanizing. Within the coming decade, China will become a

predominantly urban country; and its 100 to 150 million rural-to-urban migrants, called the “floating population”, will pose health challenges in terms of new risk factors and demands for appropriate health care arrangements. India’s urbanization is not as marked but its rural-to-urban migrants mostly join the informal economy with little social provision or social security. Gender imbalances plague both India and China, producing their overwhelming share of the world’s “missing women”. The abysmal premature loss of women, from the womb through adulthood into widowhood, exacts a huge human toll. Some have hypothesized that shortages of women may benefit them by changing the dynamics of the marriage market, but just as likely are the negative implications of harmful social practices like the expansion of commercial sex work.

Probably the biggest demographic difference between India and China in future will be population aging. China is experiencing a most dramatic fertility decline and the fastest onset of aging in human history. Aging in India will be less precipitous. By 2050, China is projected to have 432 million people, and India 330 million, over the age of 65. Population aging will obviously change the spectrum of disease threats, with an increase in the prevalence of chronic and degenerative diseases. It will also impose dependency burdens on the national economy, household finances, and family care. In some cases, aging and urbanization are interactive, as young people migrate to the cities, leaving behind the rural elderly without family structures able to manage home-based care. There may be broader macroeconomic effects as well. Bloom *et al.* (2006) have postulated that societies experience a “demographic dividend” as they shift through the demographic transition from high to low birth and death rates. Proportionately more youthful workers in a changing age structure contributed 2 per cent of the per capita income growth in East Asia during the period 1965–90, even when other factors were controlled. China reaped its demographic dividend earlier than India, which will enjoy an economic boost in the early decades of the twenty-first century.

Table 3.2 summarizes the leading causes and risk factors of death in India and China. Both countries have a high prevalence of chronic and degenerative diseases—stroke, pulmonary and heart disease, cancers and injury. Epidemiologic forecasting suggests that the burden of chronic and degenerative diseases will weigh even more heavily into the future. Both countries, however, also have significant subgroups plagued by the diseases of poverty—common infections, child undernutrition, and maternity-related risks. Such epidemiologic diversity is greater in India than China. To manage such diversity, the health care systems of both countries will need to build in flexibility to respond to local priorities.

Two epidemiologic factors are worthy of special attention. Perhaps the single health indicator most decidedly against India is the prevalence of child malnutrition at 46 per cent in comparison to China’s 8 per cent (Kumar 2007). India’s less visible chronic childhood malnutrition, in contrast to acute hunger and starvation, challenges Sen’s theory of how to address these “silent famines” (Sen and Drèze

Table 3.2. Causes of death, burden of disease, and selected risk factors (% of total deaths), India and China

Causes of death		Burden of disease		Risk factors	
INDIA					
Heart disease	14.8	Respiratory infections	8.5	Child underweight	9.8
Respiratory infections	10.7	Low birth weight	6.6	High blood pressure	9.8
Stroke	7.4	Heart disease	5.1	High cholesterol	8.2
Low birth weight	5.2	Diarrhoeal diseases	5.1	Tobacco	7.6
Chronic obstructive pulmonary disease	4.7	Depressive disorders	4.9	Low fruit and vegetable intake	5.7
Diarrhoeal diseases	4.4	Unintentional injuries	4.3	Unsafe water and sanitation	5.4
Tuberculosis	3.5	Childhood-cluster diseases	3.4	Indoor smoke from solid fuels	4.3
HIV/AIDS	3.5	HIV/AIDS	3.4	Unsafe sex	3.4
Childhood-cluster diseases	2.8	Tuberculosis	2.8	Physical inactivity	3.3
Other unintentional injuries	2.2	Stroke	2.5	Iron deficiency	2.7
CHINA					
Stroke	18.1	Stroke	7.3	High blood pressure	14.7
Chronic obstructive pulmonary disease	14.1	Depressive disorders	6.3	Tobacco	8.2
Heart disease	7.7	Perinatal conditions	5.6	Alcohol	5.4
Stomach cancer	4.5	Chronic obstructive pulmonary disease	4.6	Indoor smoke from solid fuels	5.0
Liver cancer	3.6	Road traffic accidents	3.7	Low fruit and vegetable intake	4.2
Trachea, bronchus, lung cancers	3.5	Unintentional injuries	3.3	High cholesterol	3.5
Respiratory infections	3.2	Respiratory infections	3.1	Urban outdoor air pollution	3.4
Perinatal conditions	3.0	Self-inflicted injuries	2.8	Physical inactivity	2.5
Self-inflicted injuries	3.0	Ischaemic heart disease	2.6	Overweight and obesity	2.2
Tuberculosis	3.0	Diarrhoeal diseases	2.5	Contaminated injections	2.0

Note: Figures for leading risk factors for India are extrapolated from summary data on South-east Asia.
Sources: WHO Statistics, 2004; Disease Control Priorities Project (<www.dcp2.org>); Ezzati *et al.* (2002)

Table 3.3. Health indicators, India and China

Indicator	India	China
Total expenditure on health (% of GDP), 2004	5	4.7
Public health expenditure (% of GDP), 2003	1.2	2
Private health expenditure (% of GDP), 2003	3.6	3.6
Health expenditure per capita (PPP US$), 2003	82	278
Physicians (per 100,000 people), 2004	60	106
Hospital beds (per 10,000 population), 2002 (India), 2003 (China)	7	22
Children under 5 underweight for age (%), 1999 (India), 2002 (China)	44.4	6.1
One-year-olds fully immunized against measles (%), 2004	56	84
One-year-olds fully immunized against tuberculosis (%), 2004	73	94
Births attended by skilled health personnel (%), 1996–2004	43	96
Prevalence of HIV (% age 15–49), 2005	0.9 [0.5–1.5]	0.1 [<0.2]
Prevalence of smoking (% of adults), women, 2002–04	17	43
Prevalence of smoking (% of adults), men, 2002–04	47	67

Sources: WHO (2004); UNDP (2006).

1989). That half of India's children suffer from retarded growth and development is an abysmal societal failing. On the other hand, China has an alarming rate of adult male smoking, 67 per cent, which is nearly half again as high as India's 47 per cent. This casts a long, dark epidemiologic shadow over China's future. China must mobilize its political will and implement evidence-based health policies that have been demonstrably effective against tobacco consumption. Anti-smoking interventions such as a heavy cigarette tax, ground rules on advertising, especially among children and adolescents, and no-smoking zones in workplaces have yet to begin to be implemented in China, or for that matter in India!

III.2 Reform of Health Systems

Berman (2006) has hypothesized that despite different paths, there are now many similarities in the health care systems of India and China. Both countries spend about 5–6 per cent of GDP on health (Table 3.3). Public funding is dwarfed by private financing of out-of-pocket payments for drugs and outpatient care. Rural health care is provided by unregulated, poorly trained, and unqualified for-profit rural doctors operating self-employed businesses. Citizens in both countries have limited access to health insurance and thus little financial protection against catastrophic expenditures for illness (Mahal and Yip 2007). Both have underdeveloped professional norms, standards and associations.

While these similarities are evident, the very different institutional contexts and capacities in these two countries are even more noteworthy. Given differing GDP levels, the per capita health expenditure in China, at US$278, is more than three times the expenditure in India, at US$82 (UNDP 2004). While much can

be achieved at these per capita expenditure levels, it has been noted that India's and China's health expenditures are still modest.[2] China's number of doctors per population is nearly double that of India's. India's struggle to improve the productivity of its publicly financed system, staffed by civil servants, is constrained by insufficient resources and lack of incentives and support. China, by contrast, has a polarized public sector: at one pole, publicly owned tertiary hospitals driven by for-profit incentives; at the other, lower-level public facilities that are feebly supported. China's regulatory system micro-manages health activities while India's suffers from weak management and poor enforcement. Public health in India is fragile due to lack of skilled personnel, funding and systems, while China inherited a Soviet-style public health system that has only received an infusion of public resources since the SARS epidemic. Most important of all, the leaderships of both countries must work hard to secure public confidence that current reform efforts are on the right track.

The ultimate direction of health care reform in both countries is unclear. The established models in Western Europe or North America, also embedded in market-based economies, somehow seem to be neither achievable nor affordable in India and China. Public financing of publicly operated service provision may be more equitable in theory but seems insufficiently efficient and affordable, and private systems suffer from market failures of information asymmetry, inability to contain costs, and profit-driven incentives. There is no doubt that each country will craft new health care systems shaped to its political, economic, cultural and health contexts. Although they will have components that resemble those of developed economies, each will most likely develop fresh models for the twenty-first century. Exchange, sharing and joint learning between India and China as they move forward would seem highly beneficial.

III.3 Development and Health

As Sen has so elegantly underscored with his comparative analysis of India's Kerala state, much of India's and China's health performance will depend on developments beyond medical care per se. Two powerful influences on the health status of populations are the social determinants of health and the governance of health affairs.

Basic education has highly positive health effects for individuals, families and communities. Women's empowerment will improve not only the reproductive health of women but also the health, education and economic well-being of entire families. Sen has examined both of these social determinants but one of enduring interest for him has been the relationship of health and economic performance (Sen 1999c).

[2] A senior Chinese policy-maker observed that the cost of providing anti-hypertensive medication to China's 160 million hypertensive adults would essentially consume all of the funds in the health sector.

On the basis of the experiences of India and China, Sen argues that the relationship of health to economic productivity is more nuanced than the customary view that wealth leads to health. Indeed, Sen argues that the production of wealth requires a healthy population. That is why he has emphasized that China's economic take-off was aided, in part, by the earlier universal provision of basic education and health care, which prepared the workforce to exploit economic opportunities (Sen 2005). He wryly observes that Mao might have been surprised that these social achievements under communism would eventually help China succeed in global capitalism (personal communication)! One of India's handicaps is an incompletely educated workforce, although India's economic rise was helped by an elite, English-language higher education system that produced a highly competitive skilled labor force. Health also exerts other effects on the economy. Health security and stability is an essential foundation for economic productivity, as demonstrated by the setbacks to India during the plague epidemic in 1996 and China during the SARS epidemic in 2002. In a world economy fueled by trade in goods and services, the spread of infectious diseases can paralyze the economy. Both India and China will likely experience both progress and setbacks in this area.

Sen has also emphasized that healthy people depend on sound "health governance" (Sen 1999b). He often cites a study by Sudhir Anand and Martin Ravallion (1993) showing that it is the level of public expenditure on health and education, not overall income (or GDP), that matters greatly in advancing these two public goods. Sen also highlights the importance of democratic participation. It is an under-appreciated fact that a vigorous civil society contributes to public health by complementing governments in service provision and also by operating as a "watchdog" in detecting deficiencies and advocating improvements. A case in point is the control of environmental pollution, where feedback, advocacy and even protests by those affected are important sentinels and leverage points for addressing environmental threats. India's vibrant community of non-governmental organizations provides a robust counterweight to state control and responsibility. While China's civil society and media are expanding, their role and functions are more contained than in India. China may therefore be comparatively handicapped in speedily addressing the emerging environmental and other health threats, which are better addressed through an active citizenry.

IV. Conclusion

In some ways, India and China confront similar challenges to their health sectors—demographic and epidemiologic change, health-sector reform, and the development aspects of health change. Both countries must correct for well-recognized

market failures and also policy and regulatory insufficiencies. However, India and China have very different institutional foundations for sorting out their health responses. The nature of health governance differs markedly. For some health responses, where efficient public-sector management is important, China may be more effective. For other health responses, where popular feedback and pressure are important, India's reaction may be more robust and durable.

For both India and China, the challenge may be less that of "prolonging life" and more the "art" of ensuring more equitable distribution of improvements in health (Sen 2002). Both nations have enormously diverse populations. The differences between India and China are as great as those within India and within China. These are, after all, single political entities encompassing ancient civilizations, the most populous in the world. And what happens to the "art of prolonging life" in these two countries is of undeniable importance to the future of world health.

References

Agarwal, B., Humphries, J., and Robeyns, I. (eds) (2005), *Amartya Sen's Work and Ideas: A Gender Perspective* (London: Routledge).

Anand, S., and Ravallion, M. (1993), "Human Development in Poor Countries: On the Role of Private Incomes and Public Services", *Journal of Economic Perspectives*, 7: 133–50.

Berman, P. (2006), "Unintended Convergence: Private For-Profit Health Care in Rural India and China", paper presented to the Health Systems, Population Health, and Development in China and India Workshop, New Delhi, 25 March 2006.

Bhalla, J. (1992), *Uneven Development in the Third World: A Study of China and India* (London: Macmillan, 1992).

Bloom, D. E. *et al.* (2006), "Demographic Change and Economic Growth: China and India", paper presented to the Health Systems, Population Health, and Development in China and India Workshop, New Delhi, 25 March 2006.

Blumenthal, D., and Hsiao, W. (2005), "Privatization and its Discontents", *New England Journal of Medicine*, 353: 1165–70.

Chen, L. C. (1987), "Coping with Economic Crisis: Policy Development in China and India", *Health Policy and Planning*, 2: 138–49.

—— Huq, E., and D'Souza, S. (1981), "Sex Bias in the Family Allocation of Food and Health Care in Rural Bangladesh", *Population and Development Review*, 7: 54–70.

Dreze, J., and Sen, A., (2002), *India: Development and Participation* (Oxford: Oxford University Press).

Ezzati, M., Lopez, A. D., Rodgers, A., Vander Hoorn, S., and Murray, C. J. L. (2002), "Selected Major Risk Factors and Global and Regional Burden of Disease", *Lancet*, 360: 1347–60.

Huang, Y., and Khanna, T. (2003), "Can India Overtake China?", *Foreign Policy*, July–August: 74–81.

Jamison, D. *et al.* (eds) (2006), *Disease Control Priorities in Developing Countries* (New York: Oxford University Press).

Kumar, A. K. S. (2007), "Why Are Levels of Child Malnutrition Not Improving?", *Economic and Political Weekly*, 42: 1337–45.

MacFarquhar, R. (2005), "Introduction", in E. Friedman and B. Gilley (eds), *Asia's Giants: Comparing China and India* (New York: Palgrave Macmillan).

Mahal, A., and Yip, W. (2007), "Financial Risk Protection and Access to Care: A Comparative Assessment of the Indian and Chinese Health Systems", paper presented to the Harvard China–India Development and Relations Symposium, New York City, 31 March 2007.

Malenbaum, W. (1982), "Modern Economic Growth in India and China: The Comparison Revisited, 1950–1980", *Economic Development and Cultural Change*, 31: 45–82.

Sen, A. (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Clarendon Press).

—— (1990), "More than 100 Million Women are Missing", *New York Review of Books*, 37(20): 61–6.

—— (1991), "The Kerala Difference", *New York Review of Books*, 38(17); available at <www.nybooks.com/articles/3117>.

—— (1994), "Population: Delusion and Reality", *New York Review of Books*, 41(15): 62–71.

—— (1996), "Objectivity, Health and Policy", in M. Dasgupta, L. C. Chen and T. N. Krishnan (eds), *Health, Poverty and Development in India* (New Delhi: Oxford University Press).

—— (1999a), *Development as Freedom* (Oxford: Oxford University Press).

—— (1999b), "Economic Progress and Health", paper presented at the Ninth Annual Public Health Forum, London School of Hygiene and Tropical Medicine, April 1999.

—— (1999c), "Economics and Health", *Lancet*, 354, Suppl.: SIVZO.

—— (2002), "Why Health Equity?", *Health Economics*, 11: 659–66.

—— (2004), "Passage to China", *New York Review of Books*, 51(19); available at <www.nybooks.com/articles/article-preview?article.id=17608>.

—— (2005), "Wrongs and Rights in Development", *Prospect Magazine*, October; available at <www.prospect-magazine.co.uk/vis_index.php?select_issue=374>.

—— and Drèze, J. (1989), *Hunger and Public Action* (Oxford: Clarendon Press).

—— —— (1995) *India: Economic Development and Social Opportunity* (Oxford: Clarendon Press).

UN (2005), *World Population Prospects: The 2004 Revision* (New York: United Nations).

UNDP (2006), *Human Development Report 2006* (New York: UNDP).

WHO (2004), "WHO Core Health Indicators", available at <www.who.int/whosis/database/core/core_select.cfm>.

Yusuf, S., and Winters, L. A. (eds) (2007), *Dancing with Giants: China, India and the Global Economy* (Washington, DC: World Bank).

CHAPTER 4

SUSTAINABLE HUMAN WELL-BEING

AN INTERPRETATION OF CAPABILITY ENHANCEMENT FROM A "STAKEHOLDERS AND SYSTEMS" PERSPECTIVE

KANCHAN CHOPRA

I. The Capabilities Approach, Development as Freedom and "Sustainable Freedoms"

The capabilities approach of Amartya Sen maintains that in examining and evaluating an individual's well-being, we should look not only at what he achieves for himself but also what he is "free to achieve". Since the 1980s, the capabilities

approach has grown from a position in welfare economics to a wide-ranging position in the philosophy of development, culminating in Sen's book *Development as Freedom* (Sen 1999). According to this position, all capabilities that humans could acquire are to be understood as falling within the ambit of one or the other of five freedoms: political freedoms, economic facilities, social opportunities, guarantees of transparency and protective security.

Some scholars, such as Gasper and Stavern (2006), argue, however, that freedom needs to be seen as one value among a number of other significant values such as justice, respect, friendship and care. The literature has also generated interesting discussions on whether or not a listing of capabilities is central to the operationalization of the approach. Sen (2006) maintains that the listing must always be context-specific. This is controversial. While examples of lists drawn up, sometimes indirectly, to assist in global assessments (such as the UNDP's Human Development Index) can be found, some scholars, such as Martha Nussbaum, maintain that access to certain capabilities needs to be treated as universal.

With respect to concerns regarding development and the environment, Sen has maintained that "the notion of 'sustainable freedoms' can add something substantial to the living-standard-based notion of sustainable development. It can combine the very important notion of sustainability—rightly championed by Bruntland and Solow—with a view of human beings as agents whose freedoms matter, rather than seeing people simply as patients who are no greater than their living standards" (Sen 2006: 360).

Viewing human well-being from these two parallel perspectives, namely capabilities and sustainable freedoms, this chapter interprets capability expansion in the context of achieving sustainable freedoms for identified social groups. It argues that in fast-growing economies, the "resilience and adaptive change approach", as applied dynamically to social systems and their evolution over time, provides one standard against which a listing of capabilities may make a great deal of sense. In fast-paced development, price triggers often change both the values and the vector of functionings facing individuals. Fixed and time-honored roles for individuals are changed; in other words, the parameters defining communities, households and other social entities or "systems" within which individuals live and function are changed. Expansion of individual capabilities depends on the nature of linkages between the sub-group to which the individual belongs and other parts of the social system. The question this chapter poses is: could an understanding of the concept of the "dynamic resilience" of a social system over successive cycles of development be significant in determining which capabilities of individuals are critical in ensuring that development as freedom is sustainable? A list of such capabilities treating individuals nested in subsystems whose dynamic resilience over time provides the touchstone for compiling this list, provides a policy dimension to the notion of capability. I believe it also extends the notion of "development as freedom" in the Senian tradition.

II. The Nested-System Approach and Sustaining Individual Freedoms

The expansion of individual freedoms takes place within social systems or their sub-entities which relate to each other dynamically over time. As the entities comprising the social system expand, the system as a whole expands, offering more choices to all. However, it may not expand continuously. The system contains dynamics that can be called "adaptive renewal cycles", with important feedback loops. The term "panarchy" (Holling 1986) is used to capture the dynamics of adaptive cycles that are nested. Consider, for instance, a household nested successively in a community, in a nation and in the global context. Connections between these levels ensure that economic and social systems operate in a globalized world. Changes at the global level, say in the price of a particular commodity, translate into changes in production, trade and the use of land and water at local levels. The first two stages in the adaptive renewal cycle (Holling 1986) can be both creative and conservative, thereby triggering off development processes and enlarging the freedoms of individuals nested within the system. The next stages to link this cycle to larger change and sustain the freedoms of individuals witness disturbances to the system, its reorganization, and its movement to a different equilibrium. Holling (1986) has argued that ecosystems go through regular cycles of organization, collapse and renewal. Renewal occurs as a consequence of a disturbance, initiated *do novo* or in system memory, often manifested in "feedback loops". In economics, the process resembles what Schumpeter (1950) called "creative destruction". It is a window of opportunity offered by, among other things, novelty (including new technology) and social memory. The novelty is created by the increasing capabilities and freedoms of individual components of the system, sometimes triggered by outside forces. Sustaining these freedoms requires the system to possess the adaptive and organizational capacity to proceed to newer equilibriums which sustain these freedoms.

In other words, two aspects determine the sustainability of expansion of capabilities and freedoms:

1. the inter-temporal aspects of individual freedoms and
2. the interpersonal aspects of individual freedoms.

Threshold levels of some freedoms may assume importance in order to ensure sustainability for others. Further, positive and negative shocks to the system may move it in desirable or undesirable directions. Be that as it may, evaluating the sustainability of individual freedoms requires us to position them within social systems. Social systems in turn consist of groups of stakeholders in the context of specific drivers or initiators of change.

To elucidate and substantiate the above hypothesis in greater detail, I focus in depth on a study of a development experience in the Indian state of West Bengal

(Chopra *et al.* 2005).[1] A multidimensional approach to well-being is used to define and measure changes in capabilities and associated freedoms of a set of stakeholders who benefited from the development. In addition, a methodology is developed for determining whether or not these capability expansions are sustainable.

III. Development in the Sundarbans of India: Aquaculture Exports, Human Well-Being and Sustainable Freedoms for Stakeholders

In the decade of the 1990s, increased production of shrimp in the Sunderbans region of West Bengal and its processing and export through the port of Kolkata, was driven essentially by the increasing openness of the Indian economy. This development process had wide-ranging impact on people and on the natural environment. New capabilities were created by the opening up of an export market; each agent in the market link benefited through increased capability for learning and adaptation to the dynamic situation. The region experienced an increase in economic facilities and social opportunities. However, the constraints and outcomes in each case are different for different stakeholder groups, and these differences provide insights into the manner in which well-being and sustainable freedom for different stakeholder groups can be related to each other.

Results from our study indicate that processed shrimp export generates considerable incomes in the region. Shrimp farmers increased the area devoted to aquaculture. By 2004, estimates of the area under aquaculture ranged from 42,000 to 50,000 hectares. Shrimp farmers obtain prawn seed from collectors in remote areas; collection is a labour-intensive subsistence occupation, using very little capital. A network of agents and traders transports seed to the farmer and shrimp to the processing units, who then undertake processing in accordance with international norms and supply the processed shrimp to export markets. The main stakeholders in the activity are: processing units, traders and collection agents, shrimp farmers, and prawn seed collectors.

All these stakeholders experienced an expansion of opportunities. However, growth has expanded choice in some directions while constricting it in others (see Appffel-Marglin and Marglin 1990). Prior to the changes brought about by

[1] The study was undertaken by a team from the Institute of Economic Growth, Delhi, supported by the Macro-Economics Office of the World Wildlife Foundation (WWF) and the World Bank. The members of the team were Kanchan Chopra, Pushpam Kumar, Preeti Kapuria and Nisar Ahmed Khan. See also Chopra, Kapuria and Kumar (forthcoming).

the shrimp export trade, the natural and social environment provided a set of ecosystem services to groups of people. Provision of shrimp-related livelihoods could have had either a synergistic or a trade-off relationship with the provision of these goods and services, which also contributed to human well-being, albeit through different institutions and economic activities. Such two-way linkages of ecosystems and welfare are widely acknowledged in the literature (Daly 1987; Global Biodiversity Assessment 1995; Dasgupta 2001; Millennium Ecosystem Assessment 2003, 2005; UNEP and IISD 2004). Ecosystems yield different types of consumptive and productive benefits for society through their different ecological functions. The production profile of a society, and its consumption preferences, social milieux, and cultural practices critically influence the conditions of ecosystems, which in turn affect the ability of the ecosystem to enhance well-being. While this may imply that groups of stakeholders are affected differently, this is not the main focus of the present chapter. Rather, we concentrate on the relationships between stakeholder and system characteristics which determine these differences in the longer run, thereby affecting the sustainability of their capability enhancements.

The entire process of shrimp production, processing and export in West Bengal, from prawn seed collection to farming, processing and transport to final destination, creates employment and generates income. It also provides opportunities for a large number of people by creating sources of livelihoods. We review briefly the gains accruing to different stakeholder groups located in and around Kolkata and in the hinterland extending through the regions of Minakhan, Canning and Gosaba to the hamlets around the Sunderbans. An export activity worth about Rs5,080 million annually at international prices generated Rs6,555 million in the year 2003/4, distributed as follows: shrimp farmers, 60.74%; shrimp farm workers, 13.77%; prawn seed collectors, 8.01%; transporters and retailers, 13.81%; and processing units, 3.71%.

The processing units in Kolkata employ more than 17,000 workers and generate an estimated annual income of Rs40 million. Profit for the units is between 2 to 5% of their gross value of output, amounting to Rs100 to Rs250 million. Further, in 2004 nearly 8,100 households were engaged in shrimp farming, with an average estimated yield of Rs135,324 per hectare. The total estimated income for shrimp farmers in 2004/5 is Rs5.68 billion. Since about 70% of production is exported, income generated from the export of shrimp is approximately Rs3.97 billion. Shrimp farms employ workers and buy seed from agents, who procure it in turn from post-larvae (PL) collectors. The farms then sell their output to agents and transporters to be taken to markets or to processing units.

There are about 72,000 shrimp farm workers in the Indian Sundarbans, with an estimated average annual per capita income of Rs18,000. The total annual income is therefore about Rs1.29 billion. There are around 150,000 post-larvae collectors in the Indian Sundarbans.[2] Their average annual per capita income is approximately

[2] This figure is from latest survey, carried out by Abhijit Mitra (2005).

Rs5,000. The total estimated income in 2004/5 for these PL collectors is about Rs750 million. Transport and trade margins also create incomes in the region. These are estimated on the basis of the difference in the price of shrimp at the farm gate and as purchased by the processing unit. This is Rs65 per kg on average, implying that the transport and retail margins generate an income of Rs900 million.

A shrimp farm owner earns a higher income than prawn seed collectors, farm workers and workers in the processing unit. This is true for all three regions: Minakhan, Gosaba and Canning. On average, an agricultural farm household has an income of Rs17,324, compared to Rs32,853 for a shrimp-farming household. It is therefore easy to understand the motivation for converting land from agriculture to aquaculture.

Post-larvae collectors have an annual household income of Rs24,282, with very little investment of capital. The market linkages are very good, and the capital investment needed for transport is made by the *aratdar* or agent. Therefore the PL collectors earn their income almost entirely from their own labour and the abundance of natural resources where they live. In fact, their income is higher than that of agriculture farm households and shrimp farm workers in the same region. Are they better off by other indices of well-being? Would they choose this occupation if they had other options open to them? The next section tries to answer some of these questions, using an array of indices of well-being.

IV. Human Well-Being as Enhancement of Freedom for Different Stakeholders

The limits of indices as measures of capability enhancement are well-documented. Indices are in a sense static: they can only approximate capability enhancement and augmentation of freedom. However, once the metric of expanding different freedoms is introduced into the indices framework, the horizon expands and alternative approaches to these indices emerge.

Further, keeping in mind the milieux in which different stakeholders experience enhancement of freedoms, indices are constructed to capture the changes in their capabilities. In this section, we set up and measure such indices for different stakeholder groups, such as shrimp farmers, other farmers, and PL collectors. The freedoms and the capabilities we intend to capture through these indices are the standard Senian ones, e.g. economic facilities, social opportunities, transparency guarantees and protective security.

If interest were concentrated only on economic facilities, income per capita might be the focus of attention. Extending the notion of development to include protective security reveals differences between stakeholders. The elements of protective security considered significant by diverse stakeholders differ: the most important components are identified by some as income or food security, by others as life security and health security. For agriculturists in the Sunderbans, the loss of cultivated area to floods is a major source of fluctuations in output, income and consumption. Food security is therefore inversely related to this loss. The kind of technologies used by shrimp farmers exposes them to the risk of crop failure, though current incomes may be high. For shrimp farmers, income security is approximated by the number of normal harvests in the past five years. As against this, for wage-salary earners and mixed-income households the probability of getting

Table 4.1. Measurement of indices of well-being in Canning, Minakhan and Gosaba blocks of the Sundarbans

Indicator	Measured by
Income per capita	0 to 1 scale, Rs10,000 p.a. representing 1
Income/Food security	*For agricultural farmers* 0 to 1 scale, depending on percentage of area lost to flooding *For shrimp farmers* 0 to 1 scale, the number of normal harvests in the past five years. 0 being no normal harvest in the past five years *For wage/salary earners and mixed-income households* Probability of getting work in a month (days of work in a month ÷ 30) 1 being the maximum probability *For fishermen and PL collectors* Percentage who collect fuel wood and honey along with fish for self-consumption and sale in the market *For PL collectors* Probability of PL collection in a month (days of PL collection in a month ÷ 30) 1 being the maximum probability
Health/Life security	*For all stakeholders* 0 to 1 scale, inverse of health expenditure as a percentage of total expenditure. Often based on number of working days lost or chronic ailments suffered *For fishermen and PL collectors* 0 to 1 scale, inverse of the average number of storms and accidents in the past five years, 0 being no storms and accidents and 1 being the maximum number of storms and accidents (lack of adequate weather information leads to accidents)
Absence of conflict/ Social cohesion	Inverse of conflict, which is on a 0 to 1 scale, 1 representing maximum social cohesion Nature of conflict varies: *For farmers* Conflict over title to land, in particular for shrimp farmers *For fishermen* Conflict over fishing rights *For PL collectors from Gosaba* Conflict with traders over price of PL and conflicts over forest entry for secondary collection of honey and other forest products

work determines the level of protective security they enjoy. For fishermen and PL collectors, the number of days of PL or fish collection in a month indicates the extent of income security. Further, since the Sundarbans forests also provide important items for consumption or sale, such as fuel wood and honey, the percentage who collect fuel wood and honey along with fish for self-consumption and sale in the market adds to the security of this group.

Extension of the notion of security to health and life security suggests disparate measures for different stakeholders. One proximate measure for all could be the inverse of health expenditure as a percentage of total expenditure, on a scale from 0 to 1. Often this may have to be based on the number of working days lost or chronic ailments suffered.

For specific stakeholders exposed to greater risks to health and life, such as fishermen and PL collectors, a good measure might be provided by the inverse of the average number of storms and accidents to which they have been exposed in the past five years, 0 being no storms or accidents and 1 being the maximum number of storms and accidents.

Our household survey, conducted in three blocks in the Sundarbans, Canning, Minakhan and Gosaba, covering more than 150 households, provided the information to estimate scores on the human well-being indices which throw light on dimensions of well-being other than income. Table 4.1 explains how the indices are measured from the information collected.

V. A Comparison of Indices of Human Well-Being

We have estimated the indices of human well-being in multiple dimensions for major stakeholders and for three different areas. The location of the stakeholders also determines their options, as there are significant differences in resource endowments between regions. Canning and Minakhan have a greater concentration of shrimp farms. Different technologies for shrimp farming have been tried, with a move towards polyculture with improved traditional techniques found to be best. These are also the areas in which the land has been converted from agriculture to aquaculture and then to brick kilns. Gosaba is the area with the largest concentration of fishermen and PL collectors, and where alternative experiments in livelihood have also emerged. Parts of Gosaba extend to the relatively distant and less accessible region as well.

For a comparative study across regions and stakeholders, we have selected the following four groups, in an attempt to answer a varied set of questions relevant to

the options made available to different groups, and consequently individuals within these groups:

1. *Shrimp farmers and agricultural farmers in Canning and Minakhan.* This comparison is significant in view of the change in land use from agriculture to aquaculture in these regions between 1986 and 2004. We ask what impact this change has had on human well-being.
2. *Shrimp farmers and mixed-income households in Canning and Minakhan.* This comparison assumes significance because it represents the option that might have existed for upwardly mobile agricultural farmers if the export-based expansion of aquaculture had not taken place.
3. *Post-larvae collectors and fishermen in Gosaba.* Post-larvae collection has emerged as a major source of livelihoods in recent years and it is mainly fisherfolk who have moved into this. The consequence is a change in the use of the natural resource base, at a cost of biodiversity loss. How has this made the PL collectors better off in comparison with fishermen?
4. *Post-larvae collectors and salary- and wage-earner households in Gosaba.* Salary- and wage-earner households represent the options available to households as a consequence of better educational levels in comparison to those of PL collectors. These households also illustrate the possible alternative livelihoods in the region should PL collection be banned.

V.1 Shrimp Farmers and Agricultural Farmers in Canning and Minakhan

As indicated in Table 4.2, per capita incomes for shrimp farmers in Canning and Minakhan are better than for agriculturists. However, the security of these incomes is much lower in terms of frequency of normal harvests. Also, while agricultural farmers are not exposed to any significant conflict, shrimp farmers have a relatively lower level of social cohesion due to constant conflicts. A large number of these

Table 4.2. Indicators of well-being of agricultural farmers and shrimp farmers; Canning and Minakhan

Indicator	Agricultural farmers		Shrimp farmers	
	Canning	Minakhan	Canning	Minakhan
Income per capita	0.25	0.43	0.93	0.73
Income security	1.00	1.00	0.66	0.56
Health	0.91	0.84	0.94	0.94
Absence of conflict/Social cohesion	1.00	1.00	0.50	0.50

Table 4.3. Indicators of well-being of shrimp farmers and mixed-income households; Canning and Minakhan

Indicator	Mixed-income households		Shrimp farmers		Measured by
	Canning	Minakhan	Canning	Minakhan	
Income per capita	0.56	0.58	0.93	0.73	0 to 1 scale Rs10,000 p.a. representing 1
Income security	0.70	0.75	0.66	0.56	Probability of getting work in a month, 1 being maximum probability
Health	0.92	0.91	0.94	0.94	0 to 1 scale, inverse of health expenditure as a percentage of total expenditure
Absence of conflict/ Social cohesion	0.9	1.00	0.5	0.5	Inverse of conflict, which is based on a 0 to 1 scale, 1 representing the maximum number of conflicts

conflicts are over property rights to land for ponds. On the whole, shrimp farmers in Minakhan seem to be worse off than those in Canning. For agriculturists, it is not so easy to rank the two regions.

V.2 Shrimp Farmers and Mixed-Income Households in Canning and Minakhan

Mixed-income households in Canning and Minakhan have lower levels of income compared to shrimp farmers but enjoy more security with respect to the availability of this income. They also have lower levels of conflict, since they do not have to face issues related to property rights for land. The indicators of well-being in table 4.3 illustrate this.

V.3 Post-Larvae Collectors and Fishermen in Gosaba

Gosaba is the area in which the increase in shrimp farming has led to the emergence and consolidation of post-larvae collection as a source of livelihoods. Table 4.4 verifies that when compared to fishermen, this has meant an increase in income and a decreased risk to life, since they need not venture onto the high seas. Food security, from the collection of food from bodies of water, remains unaffected. Additionally, there is a higher level of income and of food and life security. However, chronic health problems arising from the nature of the work were more severe: 85%

Table 4.4. Indicators of well-being of post-larvae collectors and fishermen, Gosaba

Indicators	PL collectors	Fishermen	Measured by
Income per capita	0.50	0.30	0 to 1 scale, Rs10,000 p.a. representing 1
Income security	0.40	0.60	Probability of number of days of collection of fish/PL in the month
Food security	0.65	0.60	Percentage of fishermen/PL collectors who collect fuel wood and honey along with fish for self-consumption and sale in the market
Life security	0.90	0.60	0 to 1 scale, inverse of the average number of storms and accidents in the past five years, 0 being no storms or accidents and 1 being the maximum number of storms and accidents
Health	0.38	0.36	0 to 1 scale, inverse of health expenditure as a percentage of total expenditure or inverse of frequency of complaints of chronic ailments
Absence of conflict/ Social cohesion	0.50	0.75	Inverse of conflict, which is based on a 0 to 1 scale, 1 representing the maximum number of conflicts

complained of these problems, though they did not incur much expense from them. Conflict was greater on several counts: 75% of PL collectors reported conflicts over price with the *dadun*, the trader who buys the catch and is often the source of credit for equipment such as the net; while 55% reported dependence on the *dadun* for the sale of their catch. *Dadun*s also lent them money for buying their net: about 87% received this credit facility, which then limited their options for sale.

V.4 Post-Larvae Collectors and Salary/Wage Earners in Gosaba

As the indicators in table 4.5 show, salary/wage earners in Gosaba have higher incomes, more security of income, fewer health problems and very little conflict as compared to post-larvae collectors. Conflicts these households can recall involve law and order problems. The salary/wage earner households have higher levels of literacy, with at least half the family literate. They travel to the mainland for work. On all indices, salary and wage earners are better off than PL collectors. This has important implications for livelihood options which can be promoted in the region.

To highlight the main findings, based on primary surveys and secondary data with respect to different agents/stakeholders:

1. Shrimp farmers in the areas of intensive conversion of land from agriculture to aquaculture are better off than agriculturists in terms of per capita incomes.

Table 4.5. Indicators of well-being of post-larvae collectors and salary/wage earners, Gosaba

Indicator	PL collectors	Salary/Wage earners	Measured by
Income per capita	0.50	0.94	0 to 1 scale, Rs10,000 p.a. representing 1
Income security	0.40	1.0	Probability of getting work in a month, 1 being maximum probability
Health	0.38	0.90	0 to 1 scale, inverse of health expenditure as a percentage of total expenditure
Absence of conflict/ Social cohesion	0.50	1.0	Inverse of conflict, which is based on a 0 to 1 scale, 1 representing the maximum number of conflicts

However, the security of these incomes is much lower in terms of frequency of normal harvests. Also, while agricultural farmers are not exposed to any conflict, shrimp farmers have a relatively lower level of social cohesion due to constant conflicts. Aquaculture farmers have to face higher conflict and more income insecurity than agricultural farmers.

2. Collectors of prawn seed from the wild have higher incomes and reduced life insecurity when compared to fishermen. However, they do not enjoy the same level of income security, health security and freedom from conflict as salary and wage earning households in the same area. Alternative livelihoods based on the latter kind of household seem a viable option if collection of seed from the wild is to be discouraged in order to prevent biodiversity erosion.

VI. From Measurement to Analysis: Towards System Analysis for Sustaining Well-Being

The different stakeholders in our study can be viewed either as individuals or as members of groups nested in the larger socio-economic system. From a welfare perspective, we can draw policy conclusions for both situations. The important questions come into view when the issue of sustainable freedoms is raised. I examine briefly the kinds of questions which arise when we look at two of the stakeholder groups: post-larvae collectors and processing industries.

The PL collectors are the least-endowed and least-privileged group, and are also at the end of the value chain initiated by the drive to increase exports. Their material well-being has improved, perhaps at the cost of social cohesion and health security. All this, however, is familiar. We know that trade-offs can exist between indices of well-being. The more significant question is: how long can even the improvement in economic facility be sustained? If thresholds of minimal social cohesion or health security essential for functioning are not obtained, individual economic benefit itself may be impaired.

The PL collectors are at the end of a chain of events triggered by a process with in-built uncertainty arising from several sources, some of which are:

- the unpredictable and complex international environment, where processing units face changing non-tariff barriers and incomplete information.
- technological and land rights issues, leading to income uncertainty for shrimp farmers.
- change in land use from agriculture and sometimes mangroves to aquaculture; as mangroves are nurseries for PL production, this feedback loop may affect the availability of wild PL itself. Even if the supply of wild PL were unlimited and continuous, the labour costs of collection may increase.
- the possible emergence of hatcheries as an alternative source of PL which could erode their income levels due to falling demand.

It is clear from these scenarios that if we need to focus on sustainable freedoms, even in relation to the limited sustainable economic opportunities for PL collectors, we need to view them as a stakeholder group nested in a larger socio-economic system. At present, this stakeholder group is part of the expansion phase of an adaptive renewal cycle and hence it experiences some of the benefits. However, a shock to the system from any of the sources listed above will harm its longer-term interests.

Consider, in contrast, the longer-run scenario for the processing factory owner. The unpredictable and complex international environment is the main source of uncertainty he needs to reckon with. And given his human and financial capital resources, he has a larger set of options. His access to information with respect to international markets also places him in a positive feedback loop with respect to advance planning for contingency.

This comparison points in two directions. It validates the significance of the "functioning and capability" approach as a better method for comparisons of well-being. But it also indicates that this approach falls short when the focus is on inter-temporal aspects of sustaining freedoms. The method to adopt in this case is to nest individuals in systems (whether social or economic) and examine the characteristics of these systems in relation to the individual. The continuing significance of terms such as "location in the adaptive cycle", "feedback loops to well-being" and "critical thresholds of functioning" in the above analysis of stakeholder groups supports this

contention. Some questions one would need to focus on and answer in this context are:

1. How does a stakeholder group with expanding short-run freedoms relate to the social and economic system?
2. Does the expansion of individual opportunity in any way exert negative or positive feedback on critical functionings of the system which interact with the sustainability of individual opportunity?
3. What minimal capability expansion is necessary for different stakeholder groups to develop the wherewithal to respond to external shocks without a loss of functioning?

In other words, ensuring sustainable freedoms for individuals located in developmental situations requires that we position them in a context of dynamic social systems. While the study of a few stakeholder groups in West Bengal illustrates this, it only provides a pointer. Developing further the capability approach to well-being, in particular its expansion to sustainable freedoms and its links to the "resilience network approach" to social systems, will throw more light on the nature of sustainable well-being.

References

Appfel-Marglin, F., and Marglin, S. (1990), *Dominating Knowledge: Development, Culture and Resistance* (Oxford: Clarendon Press).

Berkes, F., Colding, J., and Folke, C. (eds) (2003), *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change* (Cambridge: University of Cambridge Press).

Chopra, K., Kumar, P., Kapuria, P., and Khan, N. A. (2005), *Trade, Environment and Rural Poverty: A Study in the Indian Sundarbans* (Delhi: Institute of Economic Growth).

—— Kapuria, P., and Kumar, P. (forthcoming), *Biodiversity, Land-Use Change and Human Well-Being: A Study of Aquaculture in the Indian Sundarbans* (*Oxford: Oxford University Press*).

Daly, G. (1997), *Nature's Services* (Washington, DC: Island Press).

Dasgupta, P. (2001), *Environment and Human Well-Being* (Oxford: Oxford University Press).

Gasper, G., and Stavern, I. van. (2006). "Development as Freedom—And As What Else?", in B. Agarwal, J. Humphries and I. Rubyens (eds), *Capabilities, Freedom and Equality* (New Delhi: Oxford University Press).

Global Biodiversity Assessment (1995), *Global Biodiversity Assessment* (Cambridge: Cambridge University Press).

Holling, C. S. (1986), "The Resilience of Terrestrial Ecosystems: Local Surprise and Global Change", in W. C. Clark and R. E. Munn (eds), *Sustainable Development of the Biosphere* (Cambridge: Cambridge University Press).

Millennium Ecosystem Assessment (2003), *Ecosystems and Human Well-Being: A Framework for Assessment* (Washington, DC: Island Press)

—— (2005), *Ecosystems and Human Well-Being: Policy Responses. Findings of the Responses Working Group* (Washington, DC: Island Press).

Mitra, A. (2005), *Study of the Evaluation of Fin Fish Juvenile Loss due to Wild Harvest of Tiger Prawn Seeds from Coastal West Bengal* (Delhi: Institute of Economic Growth).

Nussbaum, M. (2006), "Capabilities as Fundamental Entitlements", in B. Agarwal, J. Humphries and I. Rubyens (eds), *Capabilities, Freedom and Equality* (New Delhi: Oxford University Press).

Schumpeter, J. A. (1950), *Capitalism, Socialism and Democracy* (New York: Harper and Row).

Sen, A. (1999), *Development as Freedom*. New York: Knopf Press.

—— (2006), "Capabilities, Lists and Public Reason: Continuing the Conversation", in B. Agarwal, J. Humphries and I. Rubyens (eds), *Capabilities, Freedom and Equality* (New Delhi: Oxford University Press).

UNEP and IISD (2004), *Human Well-Being, Poverty and Ecosystem Services* (Nairobi: United Nations Environment Programme and International Institute of Sustainable Development).

CHAPTER 5

HUMAN RIGHTS AND HUMAN DEVELOPMENT

SAKIKO FUKUDA-PARR

I. Introduction

One of the important contributions of Amartya Sen has been to highlight the significance of human rights in development (Sen 1981, 1984, 2005). Human rights is one of the key concepts, along with capabilities and functionings, entitlements and fundamental freedom, that Sen has used in defining the ultimate ends of development as expansion of capabilities and human freedom. The contribution of these ideas to economics and to the theory and practice of development is well known and well documented, and has given rise to a vibrant body of new research and policy debates.

However, less well known among development economists is the significance of these same ideas to the field of human rights, particularly to the new and growing work on the role of human rights in poverty reduction and development. This is a new area of work for both theorists and practitioners. Sen's capability and human development approach provides a useful conceptual framework within which human rights principles can be incorporated, because his approach defines the

The author gratefully acknowledges comments from Kaushik Basu, Shareen Hertel and Lanse Minskler, and research assistance from Mary-Anne Hoekstra, Maximillian Ashwill, Elizabeth Chiappa and Carol Messineo

ultimate purpose of development as the expansion of human freedom (Donnelly 2003; OHCHR 2004; Alston 2005). Human development (and the capability approach) and human rights thus share a common motivation (UNDP 2000). The purpose of this chapter is to draw on Sen's work to address current issues surrounding the concept and implementation of the "human rights-based approach to development" (HRBA), and to encourage further research and policy debates in this area.

HRBA is widespread among the key stakeholders in the development field (Maher and Marks 2007). It is being adopted by increasing numbers of development agents, including local and international NGOs, civil society groups, bilateral and multilateral donors, and think tanks. Increasing numbers of human rights scholars are interested in poverty reduction and development questions. Yet the implementation of HRBA remains uneven, and is still at the margins of both development and human rights. As most writers on the issue observe, the "integration" of development and human rights has proven partial and difficult (Sano 2000; Uvin 2004; Alston and Robinson 2005; Gasper 2007). Uvin's analysis of development cooperation programmes (Uvin 2004) shows that the adoption of HRBA by development cooperation agencies has been largely rhetorical and only a few have introduced new approaches to program priorities and implementation. Other evaluations of the HRBA point to mixed experiences (ODI 2006; Darrow 2007).

The truth is that many development practitioners remain highly skeptical of the idea of human rights as a central issue of development ends and means (Ingram and Freestone 2006). As Alston and Robinson (2004) point out, they continually ask "what is the value added?" (as does, for example, Kanbur 2007), to which the reply from the human rights community is that there is a need for "value change" in development economics (Eide 2006). Rights and development have been characterized by Alston in his recent article (2005) as "ships passing in the night" even though they are heading in the same direction.

These difficulties may be due to lack of communication, as Alston suggests, or disagreement about the importance of human rights. This chapter reviews the critiques to explore whether the problems are due instead to certain limitations in the concept and practice of HRBA, with the aim of identifying ways in which contradictions can be resolved and gaps filled.

II. Human Rights-Based Development: The Concept and Policy Agenda

In the late 1990s, human rights scholars and activists began to focus on global poverty. Concerned that the evolving global order was not adequately protecting

and fulfilling human rights, they began to argue that human rights are central to both the ends and means of the development process. This new attention grew out of the end of the Cold War, which opened the way for the human rights community in the West to give more attention to economic and social rights. The rapid rise of civil society and their global networks throughout the world during this time was also a major factor. This activism took place within the UN system and the UN human rights machinery as Mary Robinson, then UN High Commissioner for Human Rights, played a leading role, often singling out global poverty as the "biggest human rights challenge" of the day (UNDP 2003; Robinson 2005). Her successor, Louise Arbour, continues to take this position, characterizing "inequalities within and between countries" as the "gravest human rights challenge" (2007: p. iii).

A discourse on development emerged, launched by a 1995 publication by the Human Rights Council of Australia, *The Right Way to Development: The Human Rights Approach to Development Assistance* (HRCA 1995). While previously, human rights and development communities worked on joint agendas such as the passage of the Convention to Eliminate All Forms of Discrimination against Women (CEDAW), and the Convention on the Rights of the Child (CRC), or had tried to use development cooperation to put pressure on governments to stop human rights abuses, this initiative attempted to combine these two fields of endeavour, which had previously developed in parallel with little interaction. The human rights-based approach to development sees development and human rights as pursuing the same objectives—defined as the realization of human rights and the respect of human rights principles in the process of development. In doing so, HRBA aims to reorient development theory and practice. HRBA often refers to program approaches adopted by development cooperation agencies that have introduced new priorities and activities. But these approaches are embedded in a broader discourse. I shall use the term HRBA broadly, synonymous with "human rights in development", to refer to that discourse.

II.1 Social Origins and Motivations of HRBA

By their very nature, human rights are ethical norms that are a product of "social ethics and public reasoning", according to Sen (2006), or the history of "social learning" according to Donnelly (2001). Norms of human rights develop because people claim that certain conditions of human life are entitlements and demand that they become recognized as human rights. These norms also develop because people confront various threats to their survival as human beings (Shue 1996). The recent emergence of HRBA must be understood as a demand driven by a concern with global poverty as an affront to human freedom and dignity, but also as a matter of injustice. Unlike the economic analysis of poverty, which looks to poor economic performance, inadequate resources or inadequate policies as the cause

of the poverty of countries, HRBA is concerned with the unequal distribution of power and wealth within and between countries.

HRBA holds duty-bearers to account, demanding action to reverse these trends. It is thus a demand for alternative social arrangements, including alternative institutions of governance and economic-social policy regimes. More specifically, it is a challenge to globalization led by market liberalization and global trade regimes, and to national policies of liberalization and reforms of social protection. Additionally, HRBA endorses and supports democratization processes that enhance the accountability of government and give citizens a greater voice in governance. According to Darrow and Tomas, "ushered in during the 1990s in response to development failures of the structural adjustment era, human rights based approaches to development have proliferated in recent years" (Darrow and Tomas 2005: 471). Similarly, Nelson and Dorsay note that "The growing prominence of human rights in development discussions raises the first fundamental challenge to a market-dominated view of development that has prevailed since the 1980s" (Nelson and Dorsay 2003: 2014).

II.2 Key Elements of HRBA and its Distinctiveness

According to the OHCHR (2006: 15), HRBA is "a conceptual framework for the process of human development that is normatively based on international human rights standards and operationally directed at promoting and protecting human rights. It seeks to analyse inequalities which lie at the heart of development problems and redress discriminatory practices and unjust distributions of power that impede development progress."

The key elements of this concept can be summed up in four points:

1. The overriding concern is human freedom and dignity.
2. The realization of human rights—all rights, including economic, social, cultural, civil and political rights—by all individuals is a central objective of development.
3. Human rights principles should be part of the process of development. These include:
   a. equality and non-discrimination;
   b. participation of individuals in activities and decisions that affect their lives, and the empowerment of people;
   c. accountability of duty-bearers to promote, protect and fulfill human rights;
   d. the indivisibility and interdependence of all human rights;[1]

[1] Meaning that each and every human right is individually important and should not be traded off for another, and that the different types of rights are mutually constitutive of one another.

   e. obligations of progressive realization and non-retrogression, and immediate realization of core minimum standards.
4. International human rights norms and standards should be applied in the development process, and governments are accountable for their obligations arising from their commitment to international laws to which they are signatories.

All four of the above key elements are consistent with the priorities of the human development and capability approach to development. Human rights and human development (or capabilities) share a common motivation and commitment to human dignity and freedom (UNDP 2000). The relationship between the two is explored by both Sen (1984, 2004, 2005) and Nussbaum (1997). They are different concepts, but overlap considerably. In fact, Nussbaum argues that the two are closely related and occupy the same terrain, being concerned with improving human welfare and quality of life as objectives of public policy. Human rights and human development are not competing concepts, however; rather, they are complementary because they perform different functions. The fields of human rights and human development (and capabilities) have developed complementary sets of analytical tools and action approaches that can be deployed to meet the common goal of securing human dignity and freedom. Later sections of this chapter will return to the different functionalities of capabilities and human rights.

HRBA's policy agenda thus overlaps with the human development agenda. It places a high priority on achieving equity, meeting basic needs for survival and security but also for participation, and promoting human agency. However, HRBA brings distinct new features to the overall framework of human development, and emphasizes others more strongly.

First, HRBA is much more explicit about accountability: because human rights are explicit entitlements, they impose claims on social institutions. So HRBA draws attention to the need for institutional reforms to secure human rights, and the accountability of duty-bearers. Provisions such as education become a matter of *accountability* rather than *charity*. Second, HRBA brings in the legitimacy of universal values and the clarity of internationally legislated norms, standards and obligations.[2] Human rights are institutionally embedded in national and international legal instruments. HRBA brings legislation and litigation together as development policy tools. The third new feature is the emphasis on equality, with attention to discrimination and to unequal power structures as a source of inequality and poverty. The fourth is the emphasis on participation and the process of development, of

[2] The universality of values is disputed by some on the grounds of cultural relativism. It is argued that international human rights regimes are based on Western values and that the Universal Declaration on Human Rights was drafted and adopted when many Asian and African countries were not independent. An extensive literature on cultural relativism has now emerged that accommodates cultural differences, including works of leading human rights scholars such as Jack Donnelly, Abdullahi Ahmed An-Na'im and Amartya Sen.

people being agents of change and making improvements in their own lives. This is a related notion of "indivisibility and interdependence"; the enjoyment of some rights is necessary for the realization of others. For example, the rights to information, to freedom of speech, to association, and other civil and political rights are essential to people demanding their rights to other rights, such as education, health care and so on (Darrow and Tomas 2005).

HRBA's methods are quite distinct from those of conventional development practice. They include the conventional tools of human rights practice and new tools of development cooperation. The conventional tools mobilize the power of law and the power of ideas and include: litigation to secure remedies to individuals for violations; stronger legislation; social mobilization to demand accountability and put pressure on governments and other powerful actors such as corporations; and naming and shaming to expose violations of human rights and hold violators (duty-bearers) to account. The new tools, such as human rights education, aim to empower people to demand their rights, and to strengthen the institutions that protect human rights, such as the ombudsman's office, the justice system and the media.

III. HRBA as Development Practice: Mapping its Spread and the Resistance to it

Since the mid-1990s, when HRBA began to spread, it has become a strong discourse in international development. In the development community, many civil society groups and NGOs (including some leading national and international NGOs such as Oxfam, Care International, Save the Children and ActionAid) have started working with HRBA, some adopting it explicitly as a framework for their poverty work and emphasizing rights to social justice as their core mission. For example, Oxfam International states: "Our mission is a just world without poverty and our goal is to enable people to exercise their rights and manage their own lives" (Oxfam International 2007). Similarly, ActionAid states, "We work with local partners to fight poverty and injustice worldwide, reaching over 13 million of the poorest and most vulnerable people over the last year alone, helping them fight for and gain their rights to food, shelter, work, education, healthcare and a voice in the decisions that affect their lives" (ActionAid 2007).

Among bilateral development agencies, adoption of HRBA has varied: some are strongly committed to it in their policy and program priorities, such as the

UK Department for International Development, the Swedish International Development Agency, and the Norwegian Agency for Development, all of whom have prepared elaborate policies and guidelines; other agencies have scarcely engaged with HRBA. But all have endorsed the principle of the importance of human rights under the OECD DAC policy of 1994, while the latest consensus policy document on HRBA, adopted in January 2007—"Action-Oriented Policy Paper on Human Rights and Development"—recognizes human rights norms as "an accepted normative framework reflecting global moral and political values". It notes that "there is growing consensus on the value of human rights principles—such as participation, non-discrimination and accountability—for good and sustainable development practice. The application of these principles builds on and strengthens good and sustainable development practice, with equal attention to process and outcomes" (OECD DAC 2007: 2, 3).

Among the UN agencies, UNICEF has been a pioneer in this approach, building on their commitment to children's human rights in their involvement with the formulation and passage of the Convention on the Rights of the Child (CRC). UNICEF and other agencies, such as the United Nations Development Programme (UNDP), have been developing policy positions and program approaches since the 1990s. In the context of coordinated programming strategy at the country level, the UN system as a whole has officially incorporated HRBA as one of the "five inter-related principles that must be applied at the country level" along with gender equality, environmental sustainability, results-based management, and capacity development (UNDG 2003), and has developed a "common understanding" of how this is to be implemented.

Implementation approaches can range from: (1) adoption of the language and "rhetoric" as important ends; (2) an agenda reflecting HRBA priorities;[3] and (3) use of human rights-specific tools and methods. According to recent reviews of development agency experience (Uvin 2004; Piron 2005; OECD DAC 2007) the adoption of the language of rights is widespread, but many have also shifted program priorities, particularly in governance areas, to strengthen civil and political rights such as access to justice. HRBA has been particularly relevant in raising issues of state–citizen linkages through use of the duty bearer–right holder and accountability perspectives, and a focus on the structural roots of poverty and exclusion as an obstacle (OECD DAC 2005). The use of human rights instruments has been relevant in the areas of women's rights (drawing on CEDAW), children's rights (drawing on CRC), and indigenous peoples' rights. The DAC study concludes that there is a real danger of agencies claiming to use HRBA when all they are doing is "rhetorical repackaging" of the same programs with the same methods and agendas. This is one way of looking at the position that the World Bank took

[3] This in itself can be subdivided into human rights-based approaches, human rights mainstreaming, human rights projects, implicit human rights work, and human rights dialog.

over many years: stating that their programs were in fact promoting human rights without using those terms.[4]

Within the human rights community, there has been increasing engagement in civil society organizations of all categories with poverty and development. While issues of poverty and social and economic rights have long been a concern for human rights organizations in the South, northern NGOs not only had neglected them but were actively hostile (Roth 2001; Neier 2003). But this has been changing gradually. Amnesty International, for example, began to look at economic and social rights in 2001, and their decision to focus on poverty for the 2008 global campaign marks a watershed in this new engagement. As for the UN machinery, a wealth of important new initiatives have been taken since the 1990s. For example, Special Rapporteurs have been appointed on Extreme Poverty, on the Right to Development, the Right to Health, the Right to Food and other relevant issues; the UN High Commission has been an active advocate; and the secretariat of the OHCHR in this area has begun important work in areas such as developing guidelines for the HRBA approach to poverty reduction.

These are important innovations and show that HRBA is widely used in development practice. But HRBA is limited and has remained at the margins of both human rights and development practice. HRBA has had little impact on some of the major actors, most notably the Bretton Woods institutions and the other multilateral economic agencies, including UNCTAD, the WTO and even UNDESA. Its reach has not extended to most national governments of the Global South, with some notable exceptions such as Brazil and South Africa. And it has had little impact on the main program work of governments supported by the international community, namely the Millennium Development Goals (MDGs) and the Poverty Reduction Strategy Papers (PRSPs).

III.1 Human Rights Agendas in the Millennium Development Goals

In an article on the MDGs and human rights, Alston surveys in detail the integration of human rights into MDG processes (Alston 2005). Despite the fact that MDGs and HRBA have substantially overlapping objectives, employ similar approaches, and involve some of the same actors, human rights agendas are barely integrated into key MDG initiatives, such as the MDG monitoring reports prepared at the country level.

[4] This was a consistent line, for instance, in World Bank statements to the human rights community in the 1990s. See statements by World Bank representative Alfredo Sfeir-Younis to UN meetings in Geneva in the 1990s (Sfeir-Younis 2001). See also Wolfensohn (2005).

As for the human rights community, the importance of MDGs has been amply recognized in a number of resolutions and reports; according to Alston, "at the superficial level, the Commission has been assiduous, even admirable, in relation to the Millennium Declaration and the MDG's" (2005: 816). Yet these were only passing references, even in resolutions that dealt with issues such as housing, the right to food, and education. MDGs have not been included in the substantive work of the experts, nor of the work of the Special Rapporteurs on a variety of themes and countries. There have been only a few exceptions to this trend, including the 2004 report of the Special Rapporteur on Health, and the working group on the right to development. Alston concludes that "MDGs have not been taken on board in any sustained way" (2005: 819).

III.2 Human Rights Agendas in Poverty Reduction Strategy Papers

Similar trends are apparent in the 55 Poverty Reduction Strategy Papers I have surveyed for this chapter (23 in Asia, 24 in Africa, seven in Latin America and one in the Middle East). PRSPs set out national policy frameworks for poverty reduction prepared by countries international support for mobilizing. Thus these documents request the policy priorities of national governments which are supported by official donors.[5]

Like the human rights commission resolutions, PRSPs make ample mention of the term "human rights", and if not human rights, at least "rights". Only a handful of PRSPs (Eritrea, Ethiopia, Grenada, Guinea Bissau, Haiti, Pakistan and Tajikistan) make no mention of either term. However, in many of them, mention is scant—in 28 PRSPs it is mentioned fewer than ten times, often not appearing until well into a document totalling some 200 to 300 pages. In most, the references are general and in passing, often in the context of a general statement of a development vision, without any sections devoted to human rights or human rights institutions. In some, only civil and political rights are mentioned. In most, there is scant reference to the nature of human rights violations and challenges such as discrimination against minorities or ethnic majorities, and little analysis of poverty from the human rights perspective, such as tracing the roots of poverty to institutionalized discrimination and lack of power.

The term "participation" is used a number of times in almost all the PRSPs reviewed, mostly to refer to participation in the PRSP formulation process, a mandated methodological feature. Participation has been the focus of pressure by

[5] PRSPs are available on the website of the World Bank, at <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPRS/O,,menuPK:384207~piPk:149093~theSitePK:384201,00.html>.

advocates and supporters of HRBA in shaping PRSPs, and it is one of the aspects of PRSPs that have been scrutinized and criticized (UNDP 2003; Piron and Evans 2004; Stewart 2005). While guidelines recommend that national strategies be based on broad consultation, many of the consultation processes have been criticized as being *pro forma* meetings with NGOs, often international NGOs. Or there may be no reference to any substantive analysis or program. From the human rights perspective, real participation is about people having a say in decisions that are made and being able to hold authorities to account. Processes such as this need to be institutionalized and go beyond ad hoc meetings, and also require a carefully designed system of voice with accountability, not just voice without response. There is little reference to the broad concept of government accountability for human rights. But transparency and sometimes accountability appear in many PRSPs in relation to good governance.

Equality and equity are consistently mentioned in PRSPs, but most often only in relation to gender equality. Indeed, several countries (Pakistan, Guinea Bissau, Central African Republic, Chad) omit this issue altogether, with not even a single mention of the terms "equality", "inequality" or "equity". What is more, the terms "discrimination" and "non-discrimination"—distinctive to HRBA—are rarely used.

HRBA agendas are absent in countries where they are clearly serious issues. For example, only some countries that have experienced civil war with serious civil rights violations refer to human rights. For example, Afghanistan, Angola and Cambodia emphasize human rights as a priority in relation to national reconciliation, stability and social order, democracy and peace. No mention is made of such issues in Eritrea, Ethiopia or Haiti—countries still facing civil strife, and where lack of accountability for past abuses and ethnic discrimination is still a live issue.

Only a handful of countries have prepared PRSPs that go beyond the rhetorical use of the term "human rights" and its associated principles of equality, participation and accountability. But these exceptional cases are especially interesting documents that incorporate some elements of the human rights perspective on poverty and development. Rwanda's PRSP is framed in the HRBA perspective, starting in the introduction, which states that the fulfilment of human rights is the overriding objective of development: "The Government of Rwanda strongly believes in the right of all its people to live a life free from poverty, hardship, oppression and insecurity. Rwanda's government is committed to securing for all its citizens a full range of social, economic and political rights and to working with its people to reduce poverty and exclusion." Elements and principles of HRBA are a central factor in several strategy areas, such as good governance (the point of which is to ensure accountability and transparency), gender equality and innovative participatory methods for identifying the poor and monitoring the poverty reduction process. Inequality is defined for income and non-income spaces and receives careful attention in the analysis of poverty and as a consequence of policy options. Uganda's

PRSP also focuses more explicitly on aspects of the HRBA agenda, particularly inequality and accountability. However, it does not give much attention to inequalities related to the ethnic strife in the north of the country. Another country that has prepared its PRSP from the HRBA perspective is East Timor. This PRSP looks at the relationship between human rights and poverty and goes beyond the conventional PRSP issues and emphasizes the need for a culture of respect for human rights, especially for women, children and vulnerable groups, and the important role of the media.

These are not the only countries that have incorporated human rights into their PRSPs. Human rights are one of the three main pillars of the government in Afghanistan. Burkina Faso develops a specific approach to promoting human rights as part of the democratic governance pillar of the program. A few other countries, such as Ghana, Angola (draft PRSP) and Liberia (interim PRSP), pay significant attention to human rights in different ways.

It is not surprising that some of the PRSPs most committed to adopting HRBA are from countries emerging from war, including Rwanda, Liberia, East Timor and Cambodia. Yet it is surprising that HRBA is not an element of the strategy of post-conflict countries such as Haiti and Sierra Leone. Further research is needed on such differences in adopting HRBA, so that greater insights may be gained into the obstacles to the adoption of this approach.

IV. HRBA as Practice: Critiques, Gaps, Contradictions

Why has HRBA remained at the margins of development policy? Are the reasons political? HRBA is being championed by the international community and sometimes by national civil society, but rarely by developing country governments. As already emphasized, human rights are tools of social pressure applied by people to challenge authority structures. It is therefore not surprising that "official" government programs such as PRSPs and MDGs do not incorporate human rights agendas. Moreover, in the international context, the motion of human rights has been used to damage the reputation of developing country governments or to sanction them. So developing country governments prefer to keep the topic out of their discussion with the international community. HRBA is therefore intrinsically inconsistent with international cooperation, since it challenges the power of nation states (Darrow and Tomas 2005).

Is the resistance to HRBA due to a mindset, a legacy of Cold War thinking combined with libertarian ideology, that rejects the principle of social and economic

rights, insisting that only civil and political rights are "real"? This view is still widely held, along with the common (and misguided) understanding of human rights as divided into two categories, civil and political rights on the one hand and economic and social rights on the other, with sharp differences, such as that civil and political rights are "negative rights" and social and economic rights are "positive rights", the latter supposedly invalid because there is no legal basis for identifying violations and responsibilities. For example, Ingram and Freestone note,

> many development practitioners—including World Bank staff—typically think of human rights abuses in terms of violations of civil or political rights, actions not necessarily associated with economic development and often perceived as neutral in terms of their impact on economic growth. Many would even argue that the provision of civil and political liberties would generally follow from sustained economic growth or is even a by-product of growing prosperity, a view that has tended to dominate the World Bank's own institutional thinking about human rights. (Ingram and Freestone 2006)

This position is rooted in the writings of Bentham, who held that "natural rights" are "nonsense upon stilts", and persists in the writings of many influential philosophers and legal scholars. It is also reflected in the popular press: for example, a recent leader in *The Economist* commenting on Amnesty's campaign against global poverty was titled "Stand up for your rights: the old stuffy ones, that is; newer ones are distractions" (*Economist* 2007).

Or has communication been inadequate between the human rights and development communities? These two fields have evolved in parallel until now, and, while each has a well-established literature of conceptual, empirical, and policy work, these have rarely been shared. Human rights and development have operated through the two separate disciplines of law on the one hand and economics on the other. Is the lack of effective integration of their agendas due to failures of understanding? The literature on critiques of HRBA is limited and difficult to pin down: Uvin (2004), Darrow and Tomas (2005), Robinson (2005) and Gasper (2007) provide useful reviews. Most often, development practitioners do not challenge the claim that human rights offers normative guidance and a framework for development. But they frequently express puzzlement over whether HRBA "adds value" to the process of development, or to the analysis of policy choices. Mary Robinson (2005) lists the major criticisms of HRBA expressed by development practitioners: it is too political and adversarial; it is unrealistic, expecting reform to be effective overnight, ignoring the underlying weaknesses of state capacity and social structures that are the reasons for major human rights concerns; it is too abstract, appealing to high principle but not concerned with policy choices where trade-offs are required; it ignores time and sequence, and does not take into account the realities of the development process, which takes a long time and sometimes needs to regress in order to progress; and it is too legalistic, relying on law that does not work in the context of weak institutional frameworks.

These critiques are both conceptual and methodological. Do they arise from tensions or fundamental contradictions between human rights and development? Are there limitations to HRBA? Or are its methods under-developed since it is a new approach still in the making?

IV.1 Methodology: Legislation and Litigation

The promotion of human rights has historically used the power of law and ideas that are inherent in human rights. Stronger legislation and litigation have been effective instruments for defending and enforcing human rights in many countries. The fact that human rights are backed by law and so can be rigorously defined and enforced is in fact the distinctive strength of human rights, and gives it unique instrumental value. Since human rights are subject to binding legal obligations that can support legal claims by citizens, and to processes of accountability in international treaty bodies, they have a unique instrumental power.

One critique of HRBA is that the legal instruments are ineffective, and thus irrelevant, in developing country contexts. The use of these instruments assumes that the state is both able and willing to protect the rights of all citizens, including the poor and marginalized (Robinson 2005; Gasper 2007). The reality is that the court system everywhere (not just in developing countries) is often tipped in favor of the rich and powerful, who have greater access to it and greater means to influence and use it effectively.

Courts are admittedly particularly ineffective in developing country contexts where institutional capacities tend to be weak. Yet there is increasing use of legal instruments and the courts to defend the rights and interests of poor people is one of the new and effective ways in which human rights—and poverty reduction—is being advanced. For example, in a series of celebrated cases in India, the courts have established that housing is a necessary means to the constitutionally guaranteed right to life, giving people protection from forced evictions if no alternative housing is arranged. In many countries, international human rights law and processes are used effectively, for example in slum eviction cases in the Dominican Republic, where more than 70,000 slum dwellers were allowed to remain in their homes in defiance of a presidential decree after the UN Committee on Economic, Social and Cultural Rights condemned a planned eviction (UNDP 2000: 22). It is true that constitutional guarantees for human rights are often a dead letter, but bringing these to a test has been a means of defending human rights. In Argentina, for example, constitutional guarantees for citizens' rights to seek state protection was put to the test when patients with HIV/AIDS were denied access to health care and medication (UNDP 2000: 22).

Hence the problem of court systems that favor the rich can be overcome when poor people make alliances with public defenders, or with human rights defenders in civil society. And in the longer term, when reforms are strengthened, the judicial

system can have important consequences for the security rights of poor people, which can in turn have multiple implications for poverty reduction and development. Poor people can defend their rights to land, for example, and save themselves from falling into destitution. Poor women can defend themselves from trafficking. Such actions can have positive implications not only for the well-being of the person affected, but also for economic growth through productivity gains. Thus one of the major thrusts of HRBA as development practice has been to provide resources and other means to support programs to strengthen access to justice. The instrumental value of the legal tools of human rights for development has been demonstrated, for example, by Agarwal, who has shown that the position of women is related to land rights, and more recently to legally enforceable urban property rights to a house (Agarwal 1994; Agarwal and Panda 2007).

Nonetheless, the critique of the legal instrumentality of HRBA also reveals the need for much better understanding of the instrumentality of human rights as law. Under what conditions and in what ways does it empower poor people, and what multiplier effects does it have on society in expanding opportunities and economic growth? Under what conditions could legal enforcement of rights be counterproductive? A common critique of human rights is its pursuit of one specific right without regard to its consequences. Glendon (1991), in her critique of "rights talk" in the US, argues that excessive demands for individual rights without regard to the social context and to community has come to impoverish American political debates. In the same way, individual pursuit of rights at the expense of community can work against development goals. Disputes over resettlement—for example, over individual rights to livelihood versus the "development objectives" of expanding power supply and irrigation—reflect such tensions between individual rights and development. These points will be taken up in section IV.3.

IV.2 Methodology: Naming and Shaming

Investigating abuses and using public exposure to shame the offender and effect behavior change has historically been the principal tool of human rights activism. However, naming and shaming can be counterproductive in the context of development cooperation. Naming and shaming assumes that the realization of rights is a matter of will and can be achieved overnight. It does not take account of the fact that realizing rights requires institutions, capacity and resources (UNDP 2000: 7). For example, the right to education cannot be achieved overnight, but requires the building of schools and the training of teachers. This applies not only to economic and social rights but also to civil and political rights. Access to justice requires the building of courts, as well as the training of lawyers and judges.

Naming and shaming by international governments and NGOs also ignores the role of geopolitics in international dialog on human rights. The use of human rights finger-pointing is selective, most often targeted against developing countries

and their governments by the powerful industrialized countries, led by the US and Western Europe, and often motivated by political or economic interests. Well-meaning international human rights activism in civil society can become a tool of geopolitical maneuvers. Naming and shaming is more likely to be effective when civil society targets their own governments, or corporations (UNDP 2000: 90, 113).

Naming and shaming is directly contrary to the methodology of development cooperation and is quintessentially about going beyond finding problems to finding solutions. While human rights successes have been built on mobilization of social movements, changes in law, and legal remedies for individual cases, development builds on policy change, building institutions and building infrastructure which opens up economic, social and political opportunities. The professional, analytical work of development is based on empirical analysis and assessment of policy options.

In a well-known article, Kenneth Roth, the president of Human Rights Watch, argues that the strength of international human rights organizations like his own is in the methodology of naming and shaming to effect behavior change (Roth 2004), although this methodology is only appropriate in cases of arbitrary discrimination and disregard for human rights, where there is a clear violation, violated, and violator. He also argues that many of the human rights issues related to poverty, and to social and economic rights, do not fall into this category, because they are "social justice" issues, often related to economic policy issues, where the "right" policy is a matter of debate. It is hard to argue, for instance, that a person's right to education or food has been compromised or violated as a result of a particular economic policy. He therefore concludes that engagement of international human rights organizations with these issues should remain limited. He rightly argues that advocacy needs to go beyond exposing the denial of rights, and that unless an effective remedy can be advocated, public outrage at widespread poverty is mere "sloganeering". He further argues that while the public might be outraged by widespread poverty—such as when a child dies from lack of health care—they would have "no idea whom to blame", or that "the blame is dispersed among a wide variety of actors. In such cases of diffuse responsibility, the stigma attached to any person, government, or institution is lessened, and with it the power that international human rights organizations can have to effect change. Similarly, stigma weakens even in the case of a single violator if the remedy to a violation—what the government should do to correct it—is unclear" (Roth 2004: 67).

Persuasive as Roth's arguments are, they also pinpoint a major weakness in current human rights methodology. The problem is not that it is impossible to identify the violator and remedy, but that HRBA has not yet developed the tools to do so. For example, in the case of a child who dies from a lack of health care, it is true that human rights specialists—mostly lawyers, political scientists, philosophers and social activists—do not know much about alternative approaches to expanding health care. But they can work with other specialists to identify key policy variables

and alternatives, and hold to account those who could do more, and exert pressure for policy reform. In fact, strong international protection of intellectual property has kept prices of life-saving medicines such as AIDS retrovirals inaccessibly high. Pressure from some governments of the Global South and from development NGOs and think tanks, using a combination of solid technical analysis of policy alternatives and targeted name-and-shame tactics, quite successfully forced governments and pharmaceutical companies to shift policy on the TRIPS agreement on intellectual property and drug pricing.

Some human rights organizations are beginning to engage with policy analysis and alternatives but this practice remains limited. So HRBA naming and shaming too often exposes the denial of a human right—"child dies of treatable illness"—but stops short of identifying the real violation and the violator, which is inadequate action on the part of of the duty-bearer to respect, protect and fulfil its obligations regarding the child's right to health care and to life.

International NGOs such as Human Rights Watch and Amnesty International are powerful forces in global governance. If they are to play a role in addressing what the UN High Commissioner calls the "gravest challenge" to human rights, they need to invest in new partnerships with the development community's policy analysts to advocate reforms in the economic, social and governance practices that constitute violations.

IV.3 Conceptual Issues: The Indivisibility of Rights

The principles of the indivisibility and interrelatedness of human rights are a cornerstone of HRBA, but are also the most problematic. They are a source of the major critique that HRBA is unrealistic and abstract and therefore offers little practical value in development planning for the practitioner. The principle of indivisibility holds that all human rights are necessary for a life of freedom and dignity, each one being individually valuable and none being more important than another.[6] It rejects the notion of hierarchy among rights, and any basis for considering economic and social rights as less important than civil and political rights. Furthermore, it is argued that human rights are "interdependent", meaning that these rights are consistent and enjoyment of one right facilitates—or makes possible—the enjoyment of another right. For example, without the enjoyment of the right to life, no other rights can be accessible; without the enjoyment of the right to education, the right to work will be curtailed. From the point of view of human development practice, it means that the promotion of human rights should not be subject to hierarchy, but should be pursued simultaneously, in a "holistic" manner.

[6] The Universal Declaration of Human Rights makes clear that all rights are equally important. The concept of "indivisibility" is affirmed in the 1986 Declaration of the Right to Development and the 1993 Vienna Declaration and Programme of Action.

For the development practitioner, particularly economists, this immediately raises the question of priorities in allocating resources and effort, since not all problems can be tackled at once. It has long been recognized among human rights theorists that the realization of economic and social human rights is subject to resource constraints. It is increasingly acknowledged that civil and political rights are also subject to resource constraints. For example, adequate institutions and infrastructure are required to vote as much as to access health care. In both cases, financial, human and other material resources are required for investment in these facilities as well as for running them. The central task of development planning and policy-making is allocating resources among competing priorities. The HRBA's principles of indivisibility would thus seem to close off all possibility for such allocation, particularly when combined with an absolutist adherence to human rights norms for each and every right of each and every individual as non-negotiable. HRBA becomes abstract and unrealistic.

Human rights theorists and institutions have responded to this issue with the principle of "progressive realization"—that state parties' obligation to realize economic and social rights must take account of historical reality and institutional and resource constraints. But this is hardly a solution, since more recently the Committee on Economic, Social and Cultural Rights specified that states have the obligation to realize *immediately*, rather than progressively, essential minimum standards and to remove discrimination, as well as an obligation of non-retrogression. All of these three obligations ignore the same institutional and resource constraints. Another, related proposal has been to develop a concept of certain development activities as entitlements (Andreassen 1997; Sano 2000). But this faces the same problem as minimum standards.

The difference between essential minimum standards and human rights norms is difficult to interpret. The reality is that in most of the sixty or so low-human-development countries of the world, large proportions of people are in poverty, measured either by the monetary measure of $1 a day or by capability deprivation as in the Human Poverty Index. The HPI values range from 30 to 65% for countries (UNDP 2005), with available data showing that a third to two-thirds of the population do not enjoy basic subsistence rights (Shue 1980). In many of these countries recent progress has been very slow, in part because economic growth has been stagnant (UNDP 2003; World Bank 2007), demonstrating that immediate realization of minimum standards is not a realistic goal. The fact that at the global level the world is far from achieving the Millennium Development Goals also shows that meeting essential minimum standards immediately is unrealistic.

Even the removal of discrimination requires resources and institutional reforms; for example, removing discrimination against girls' education in many contexts requires training female teachers, installing adequate toilets for girls, and public debates about the value of girls' education and equal rights of women to shift the

mindsets of community leaders and parents. And the principle of non-retrogression denies any justification for leaving individuals worse off.

In fact, as reviews by Uvin (2004) and Gasper (2007) show, the human rights community has been "averse to admitting conflicts of rights", and has not done enough to address this issue. Leading scholars and practitioners such as Alston and Robinson urge the community to engage more in addressing the issue (2005). If this were to be pursued, it would be useful to make two kinds of distinctions. The first would be to distinguish between prioritizing resource use rather than prioritizing rights, and the second would be to distinguish between conflicting rights and their consequences.

Prioritizing Rights or Resource Use

The principle of indivisibility and interdependence contradicts prioritizing of the rights of individuals, but it does not contradict the need to prioritize resource use. As Darrow and Tomas emphasize, the indivisibility of rights "does not mean that all rights must be elevated to the same level of priority in programmatic terms. Resource and institutional constraints will often require the prioritization and realization of different rights for the purposes of policy choice" (Darrow and Tomas 2005: 503).

The attention of HRBA should therefore focus on empirical analysis of resource allocations in national budgets, in donor budgets, in NGO budgets, and so on. This analysis will need to use the tools and information base of development economics to analyze and criticize government policy positions. The work currently being done on critical analysis of budgets from the point of view of children's rights, or of women's rights (gender budgeting) is an example of how HRBA can address resource issues. But such approaches can be expanded and applied to other areas that are relevant to the promotion of human rights. For the economist seeking the "added value" of HRBA, such exercises will show that HRBA raises new questions that policy-makers will need to answer.

Yet another path would be to work further on the state–citizen relationship and the state's capacity to meet its human rights obligations. Human rights are interdependent not only from the rights-holder perspective but also from the duty-bearer perspective. Just as an individual cannot enjoy a life of dignity and freedom with only a piecemeal realization of rights, a state is unlikely to commit itself to its human rights obligations piecemeal. The nature of the state and the state–citizen relationship is therefore fundamental to the realization of rights. But the institutional and financial capacity of a state is another condition of being able to protect, respect and fulfill the rights of citizens. HRBA needs to focus more on expanding the capacity of the state to meet its obligations, and expanding the resource base of the state. Here again, economic analysis will be needed, for example on choices in taxation policy.

However, another view sees human rights not as inherently a tool for resource allocation but as a tool for political action to challenge "elite capture" in an environment of power asymmetries (Darrow and Tomas 2005: 471–2). Ultimately, there are no "technocratic" answers to priorities that societies might set. Economics helps analyze the efficiency of resource use, but the ultimate decision depends on how the benefits are evaluated. Human rights bring ethical values into evaluating the benefits. The instrumentality of human rights is to resolve competing claims through the courts (Darrow and Tomas 2005). Another way would be through democratic, participatory consultation processes (Sen 2004; Osmani 2005).

Conflicting Rights or Consequences of Policy

The principle of indivisibility and interdependence is consistent with recognizing that action to fulfill one right can conflict with the enjoyment of another right or the rights of other individuals. For example, as Basu (1999) and others have argued, a policy to prohibit child labor, such as trade sanctions on export products manufactured with child labor, would drive out children from, say, carpet weaving for exports, but the consequence would be that the child and family suffer hunger and other privations, or that the child is sent to even more exploitative work such as prostitution. Critics say that HRBA does not recognize the need for sequencing. In fact, such policy positions are never a simple matter of trade-offs, but policy choices may be found that would achieve the most protection of the child's human rights.

The principle of indivisibility and interdependence would see all enjoyment of human rights as having a positive impact for the duty-bearer. But from the duty-bearer's perspective, policy action to protect and fulfill one right may have negative consequences for other rights. Precisely because of the complex interrelatedness of human rights in human lives, the repercussions are important. Investigating consequences, then, is critical to the further strengthening of HRBA. In such cases, it is also important to clarify and distinguish between what can be ethically acceptable to "trade away" and what cannot (see Basu 2007).

V. Human Rights and Development: The Way Forward

This chapter has shown that there are no inherent contradictions between human rights principles and development. Human rights principles, for example, do not contradict the demands of prioritization.

The absolutist interpretation of human rights principles of indivisibility is simplistic and does not recognize the complexity of the concept nor of the contexts in which the principle is applied. This leads to a blind adherence to principles that amounts to little more than sloganeering, and when combined with the passionate zeal of the human rights defender, detracts from rather than adds to the credibility of HRBA. Indivisibility does reject the ranking of rights in a hierarchy of importance. But that is not inconsistent with prioritization of resources. The concrete challenge is to develop the tools of analysis and advocacy that challenge resource use which is inconsistent with human rights priorities.

There are major limitations in the toolkit of HRBA, which stops short of identifying violations that lie in policies—economic, social and governance—which reflect the shortcomings of human rights obligations. The human rights field, dominated by law, philosophy and political science, has not engaged with real options for policy. The development field, on the other hand, dominated by economists and technical specialists, focuses on the analysis of policy options. Studies and debates are lively and ongoing. But these studies and debates are not connected to the normative priorities that human rights frameworks identify. It is often thought that economists have the "right answers" to policy choice and that human rights has little to add. In fact, there are no "right policy choices". Economic and technical analysis can evaluate policy alternatives, pointing out which ones are most economically efficient, or technically effective. But the choice of policy depends on ethical choices over such trade-offs as that between disincentives to pharmaceutical companies to invest in innovation and the expansion of access to life-saving drugs today. It is the role of human rights to guide such policy choices.

But conventional analysis of policy options is not being done with human rights norms in mind. Yet all development policies—economic, social, governance—have huge implications for the fulfillment of human rights. Macroeconomic policies such as taxation have an enormous impact on human rights outcomes, and the choices that governments make in this area determine whether government is doing enough to respect, protect and fulfill human rights (Balakrishnan 2005). Recent work, such as the analysis of budgets and their impact on child rights, and an analysis of economic policies and state obligations under CEDAW (Elson 2006), are important examples of policy analysis for human rights. These are important beginnings in developing new tools.

Naming and shaming, and social mobilization, are distinctive tools of human rights. Human rights have a unique capacity for empowerment. But unless human rights advocacy is combined with a more careful analysis of the policy options and the identification of the real violation, it will be merely rhetorical.

HRBA and the project to integrate human rights and development into a single undertaking, with common motivations and objectives, complementary and coherent methods and strategies, is still in the making. While critics from within the human rights field such as Roth (2004) propose a highly selective approach to

invoking human rights concepts and to activism on poverty issues, others, such as Alston (2004) and Alston and Robinson (2005), call for greater engagement. Still others, such as Sano (2001), have pointed out that the two fields cannot be fully integrated. The shortcomings of HRBA identified in this chapter point to a need for greater integration. As Nussbaum explains, in comparing capabilities with human rights, we need the concept of capabilities and functionings to evaluate how societies are faring, to be able to make comparisons between the past and the present, between countries, and so on. The role of human rights, on the other hand, has a different focus: to help develop ethical norms that guide designs on institutional arrangements.[7]

The merger of human development and human rights as concepts is not a question but a fact. The idea of entitlement and human rights is an integral part of the concept of capabilities (Sen 1984). While the two concepts are not the same, their interconnection goes deeper than simply the sharing of broad visions. As noted in the 2000 *Human Development Report*, "Human rights and human development are close enough in motivation and concern to be congruous and compatible, and they are different enough in strategy and design to supplement each other fruitfully. A more integrated approach can thus bring significant rewards, and facilitate in practical ways the shared attempts to advance the dignity, well-being and freedom of individuals in general" (UNDP 2000: 19).

References

ActionAid (2007), "ActionAid—About Us", <www.actionaid.org/main.aspx?pageID=2>.

Agarwal, Bina (1994), *A Field of Her Own: Gender and Land Rights in South Asia* (Cambridge: Cambridge University Press).

—— and Panda, Pradeep (2007), "Toward Freedom from Domestic Violence: The Neglected Obvious", *Journal of Human Development*, 8: 359–88.

Alston, Philip (2005), "Ships Passing in the Night: The Current State of the Human Rights and Development Debate Seen through the Lens of the Millennium Development Goals", *Human Rights Quarterly*, 27: 755–829.

—— and Robinson, Mary (2005), *Human Rights and Development: Towards Mutual Reinforcement* (New York: Oxford University Press).

Amnesty International (2005), *Human Rights for Human Dignity: A Primer on Economic, Social and Cultural Rights*; available at <web.amnesty.org/library/Index/ENGPOL340092005?open&of=ENG-AGO>.

Andreassen, Bård-Anders, *et al.* (1988), "Assessing Human Rights Performance in Developing Countries: The Case for a Minimal Threshold Approach to Economic and Social Rights in Developing Countries", in Bård-Anders Andreassen and Asbjrn Eide (eds),

[7] See Nussbaum (1997) for an exploration of the relationship between human rights and capabilities.

Human Rights in Developing Countries 1987/88: A Yearbook on Human Rights in Countries Receiving Nordic Aid.

Balakrishnan, Radhika (2005), *Why MES with Human Rights? Integrating Macro Economic Strategies with Human Rights* (New York: Carnegie Council).

Basu, Kaushik (1999), "Child Labour: Causes, Consequences and Cure, with Remarks on International Labour Standards", *Journal of Economic Literature*, 37: 108–19.

—— (2003), "Global Labour Standards and Local Freedoms", WIDER Annual Lecture 7; available at <www.wider.unu.edu/publications/annual-lectures/annual-lecture-2003.pdf>.

—— (2007), "Human Rights as Instruments of Emancipation and Economic Development", in Shareen Hertel and Lanse Minkler (eds), *Economic Rights: Conceptual, Measurement and Policy Issues* (New York: Cambridge University Press).

Darrow, Mac (2007), "A Human Rights-Based Approach to Development: Theoretical and Operational Implications for the World Bank", in Caroline Sage and Michael Woolcock (eds), *World Bank Legal Review*, vol. 2, *Law, Equity and Development* (Leiden: Martinus Nijhoff).

—— and Tomas, Amparo (2005), "Power, Capture and Conflict: A Call for Human Rights Accountability in Development Cooperation", *Human Rights Quarterly*, 27: 471–538.

Donnelly, Jack (2006), *Universal Human Rights in Theory and Practice*, 3rd edn (Boulder, Colo.: Westview Press).

Economist (2007), "Stand up for your rights: the old stuffy ones, that is; newer ones are distractions", 24 March.

Eide, Asbjorn (2006), "Human Rights Based Development in an Age of Economic Globalization", in Bard Andreassen and Stephen Marks (eds), *Development as a Human Right* (Boston: Harvard School of Public Health).

Elson, Diane (ed.) (2006), *Budgeting for Women's Rights: Monitoring Government Budgets for Compliance with CEDAW* (UNIFEM).

Gasper, Des (2007), "Human Rights, Human Needs, Human Development, Human Security: Relationships between Four International 'Human' Discourses", *Forum for Development Studies*, 1: 9–42.

Glendon, Mary Anne (1991) *Rights Talk: The Impoverishment of Political Discourse* (New York: Free Press).

HRCA (Human Rights Council of Australia) (1995), *The Right Way to Development: Human Rights Approach to Development Assistance* (Sydney: HRCA).

Ingram, Joe, and Freestone, David (2006), "Human Rights and Development", *Development Outreach*, October 2006; available at <www1.worldbank.org/devoutreach/october06/article.asp?id=378>.

Kanbur, Ravi (2007), "Attacking Poverty: What is the Value Added of a Human Rights Approach?", available at <www.people.cornell.edu/pages/sk145>.

Marks, Stephen, and Mahal, Ajay (2007), "Goals and Instruments of Poverty Reduction: Economic and Human Rights Perspectives on Children's Rights and Development Strategies", unpub. paper commissioned by UNICEF.

Neier, Aryeh (2003), *Taking Liberties: Four Decades in the Struggle for Rights* (New York: Public Affairs).

Nelson, Paul, and Dorsey, Ellen (2003), "At the Nexus of Human Rights and Development: New Methods and Strategies of Global NGOs", *World Development*, 31: 2013–26.

Nussbaum, Martha (1997), "Capabilities and Human Rights", *Fordham Law Review*, 66: 273–300; repr. in Patrick Hayden (ed.), *The Philosophy of Human Rights* (St Paul: Paragon House).

ODI (2006), "Human Rights and Poverty Reduction: Realities, Controversies and Strategies", available at <www.odi.org.uk/rights/Publications/Rights%20Meeting%20Series%20Publication%202006/intro_screen.pdf>.

OEDC DAC (2007), "Action-Oriented Paper on Human Rights and Development", DCD/DAC(2007)15 (Paris: OECD).

OHCHR (2004), *Human Rights and Poverty Reduction: A Conceptual Framework*, available at <http://ohchr.org/english/issues/povertyguidelines.htm>.

—— (2006), "Frequently Asked Questions on a Human Rights-Based Approach to Development Cooperation", available at <http://ohchr.org>.

—— (2007), "Mary Robinson, United Nations High Commissioner for Human Rights (1997–2002)", available at <www.unhchr.ch/html/hchr/unhc.htm>.

Osmani, Siddiqur (2004), "Human Rights: Building on the Capability Approach", *Journal of Human Development*, 6: 205–19.

Oxfam International (2007), "Mission Statement", available at <www.oxfam.org/en/about/mission>.

Piron, Laure-Helène (2005), "Integrating Human Rights into Development: A Synthesis of Donor Approaches and Experiences", available at <www.odi.org>.

—— and Evans, Alison (2004), "Politics and PRSP: Synthesis Paper" (London: ODI).

Robinson, Mary (2005), "What Rights can Add to Good Development Practice", in Alston and Robinson (2005).

Roth, Kenneth (2004), "Defending Economic, Social and Cultural Rights: Practical Issues faced by an International Human Rights Organization", *Human Rights Quarterly*, 26: 63–73.

Sano, Hans-Otto (2000), "Development and Human Rights: The Necessary, but Partial Integration of Human Rights and Development", *Human Rights Quarterly*, 22: 734–52.

Sen, A. K. (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Clarendon Press).

Sen, Amartya (1984), *Resources, Values and Development* (Cambridge, Mass.: Harvard University Press).

—— (1989), "Development as Capability Expansion", *Journal of Development Planning*, 19: 41–58; repr. in Sakiko Fukuda-Parr and Shiva Kumar (eds), *Readings in Human Development* (New Delhi: Oxford University Press, 2003).

—— (1999), *Development as Freedom* (New York: Alfred Knopf).

—— (2004), "Elements of a Theory of Human Rights", *Philosophy and Public Affairs*, 34: 315–56.

—— (2005), "Human Rights and Capabilities", *Journal of Human Development*, 6: 151–66.

—— (2006), "Human Rights and Development", in Bård-Anders Andreassen and Stephen Marks (eds), *Development as a Human Right: Legal, Political, and Economic Dimensions* (Cambridge, Mass.: Harvard University Press).

Sfeir-Younis, Alfredo (2001), *Economic, Social and Cultural Rights and Development Strategies* (Geneva: UNESCO Committee on Economic, Social and Cultural Rights).

Shue, Henry (1996), *Basic Rights: Subsistence, Affluence and U.S. Foreign Policy*, 2nd edn (Chichester: Princeton University Press).

Stewart, Frances, and Wang, Michael (2005), "Poverty Reduction Papers within the Human Rights Perspective", in Alston and Robinson (2005).

UNDG (2003), "*UN Statement of Common Understanding on Human Rights-Based Approach to Development Cooperation and Programming*; available at <www.undg.org/?P=221>.

UNDP (2000), *Human Development Report: Human Rights* (New York: Oxford University Press).

—— (2003), *Evaluation of UNDP's Role in the PRSP Process* (New York: UNDP).

Uvin, Peter (2004), *Human Rights and Development* (Bloomfield, Conn.: Kumarian Press).

Wolfensohn, James D. (2005), "Some Reflections on Human Rights and Development", in Alston and Robinson (2005).

World Bank (2007), *Global Monitoring Reports: Confronting the Challenges of Gender and Fragile States* (Washington, DC: World Bank).

CHAPTER 6

ENTITLEMENTS AND CAPABILITIES

YOUNG PEOPLE IN POST-INDUSTRIAL WALES

JOCELYN KYNCH

I. A Personal Note

The invitation to write a chapter for Amartya Sen's festschrift is a delight and an honor. I was lucky enough to be Amartya's research assistant in the 1980s, when he was developing his conceptualizations of famines, poverty and gender (Kynch and Sen 1983), and he has remained a key influence, his writings occupying a shelf below Marx, and consulted more often. He also encouraged me to carry out fieldwork on nutrition in Palanpur (Lanjouw and Stern 1998: 409–28) and agricultural wages in West Bengal (an influence on Pal and Kynch 2000), and I was privileged to meet his colleagues there.

I have also benefited from his interest in research beyond economics, including research into areas of natural science such as nutrition and medicine. Crossing disciplines has served me well over the years, most recently when seconded as a government social researcher, negotiating between policy-makers in health, social policy and social justice. This was in our devolved Welsh Assembly, where a report about supporting youth bore the title *Extending Entitlement*. This report sounds (and sometimes reads) like Sen, but why? For this chapter, I set out to discover if

the commissioners or authors had been influenced by Sen, and, following denials of direct influence, my search expanded into a question of how theory might be reflected in policy, albeit with mutated concepts.

II. Introduction

In this chapter, I shall consider how the devolved government of Wales has formed a distinctive approach to policy within the United Kingdom, in particular with respect to young people, and will consider to what extent it echoes Sen's body of work on entitlements and capabilities, or how Sen's framework might apply to it.

A major principle on which all devolved Welsh policy is based has been articulated by the First Minister as *progressive universalism* (Morgan 2006). According to Drakeford (2007: 175), progressive universalism in Wales is the provision of services universally in preference to means-tested or targeted services, and the "progressive part of universalism comes in providing ... additional help for those who need it most", so that "Universal services, with a progressive twist, combine the advantages of the classic welfare state with some of the benefits which can be claimed for targeting. What the approach provides, however, is a universal obligation, as well as a set of universal rights. In this model, social justice is everybody's business."

The argument I wish to make is that social policies aimed at young people in Wales, building on politically nurtured historical roots in which collective responsibility played a major part, have had a positive influence, as evidenced by outcomes for young people. Nevertheless, I shall suggest that there are some conceptual gaps, which can be addressed by Sen's work. For example, social relations involving young people could usefully be interpreted as *cooperative conflicts*, requiring attention to bargaining in place of appeals to cooperation or sympathy, if Welsh policy is genuinely to encompass its aims for social justice.

Section III briefly introduces devolved Wales, a minority interest even within UK academia. Section IV describes the evolution of social policies towards young people, and their emphasis on including young people's voices. The origins of the influential Welsh document *Extending Entitlement* are explored, and ways in which it mirrors Sen's framework of *entitlements*, *realized functionings* and *capabilities*. The fifth section concerns bi-national policy, and how it has been adapted for the Welsh policy context.[1] In section VI, drawing on my own empirical research, adult–youth

[1] In this chapter *national* refers to Wales and *bi-national* to England and Wales. "The UK" refers to all of England, Scotland, Wales and Northern Ireland.

relations in public spaces are examined as cooperative conflicts. The seventh and final section draws some conclusions.

III. A Brief Note about Devolved Wales

Wales is a small nation of about 3 million people, of whom nearly half are concentrated in the south-east, where the post-industrial Valleys cut southwards through the hills towards the capital Cardiff (once the largest coal port in the world). Arguably, Wales was the cradle of the Industrial Revolution, and also first to move away from the heavy industries associated with it (Morgan 1995; see CWM n.d. for a comprehensive bibliography). This history has helped shape attitudes to social justice: for example Drakeford suggests as influences on these attitudes a flatter class structure (fewer middle-class, more working-class people), economic structure (the lack of a major financial sector), and political structure. He claims that "Wales is the only part of the UK to have a genuinely long-term commitment to left of centre redistributive politics, stretching for more than 150 years, from 19th-century Liberalism through 20th-century Labourism and on to the present day" (Drakeford 2007: 174).

Devolution in the United Kingdom was taken forward by the Labour government elected in 1997, following 18 years of Conservative rule. A referendum in Wales produced only a slim majority in favor of the Assembly, which was set up in 1999 initially as a corporate body without any tax-raising or primary legislative powers. The Government of Wales Act 2006 separated the powers of the executive and Assembly Members and now provides for limited law-making powers on devolved matters, but not tax-raising.

The lack of economic levers may have benefited social policy, in that imaginative leverage has gone into the devolved areas, such as education, health, housing and local government, and has even seeped over into the implementation of economic, criminal justice and other policies decided bi-nationally by the UK government. However, a problem for any distinct devolved social policy can be that rhetoric does not stop at the border even when the powers do.

IV. *Extending Entitlement*: An Example of Devolved Policy

In the late twentieth century, prospects in South Wales were distressingly poor outside the prosperous parts of the capital city, Cardiff, and the Vale of Glamorgan:

the region's poverty was recognized by European Union Objective 1 special funding to most of the Valleys. Debates in Wales about poverty, as in other parts of the UK, were strongly influenced by the concept of social exclusion (Room 1995: 4–9), with an Anglo-American empiricist twist. While the welfare state was being criticized, and dismantled, solutions—best practice—were being trialed in regeneration areas, often on the council housing estates built in the 1960s. Along with regeneration came a wave of global mantras of consultation and participation (Cooke and Kothari 2001) and later, emphasis on partnerships, service-level agreements with the voluntary sector, and the relating of public finances to actual activity.

The scale of the new Welsh Assembly's flagship program for regeneration, Communities First, was ambitious—initially the 100 most deprived communities (compared with only 39 communities under the equivalent program in England). Concern was widespread, not only about area-based poverty, but also about children and young people, especially the high proportion in low-income households (36% of Welsh children in 1996/7), and especially after a report showed that in the Valleys about a fifth of young people leaving school went off the official radar: they were not in employment, training or further education, and qualitative research described their situation as *status zero* (Istance and Williamson 1996; Williamson 1997). Indeed, Communities First has often delivered child and youth provision because that has been identified as the communities' own main priority.

A combination of research and activist interests and events drove forward an agenda on children and young people, one result of which was *Extending Entitlement: Supporting Young People in Wales* (NAW 2000). A selection of the main actors has told me that they drew directly on the social exclusion literature and either the rights-based literature or their skilled observations.

Interestingly, the early debates about youth, about the running down of youth services, which had not been protected from spending cuts in the late twentieth century, and about problems of young people whose access to state benefits had been removed, involved future Assembly Members and ministers as well as UK ministers, who shared strong views on equality and rights. Many of these were women. (The remarkable achievement in the 2003 elections of returning 50% women Assembly Members in a nation dominated by elderly white male councilors followed the Welsh Labour Party's affirmative action in 1999: see Beddoe 2004 and the literature cited there.) It is possible that women's achieving their political potential has been another lever on policy.

Politicians are also moved by events, and the coinciding of these debates with the uncovering of child sexual abuse in children's homes (resulting in the Waterhouse Tribunal's report *Lost in Care*, DOH 2000), facilitated the appointment of an independent Children's Commissioner in 2001. Furthermore, special provision for Wales was made in the UK Learning and Skills Bill 2000 (NAW 2000: annex 2), which expressed a need for study by an Expert Reference Group to recommend how best to support young people in Wales, especially the marginalized. In Wales,

a young person is defined as aged up to 25 years, which has enabled support to young people who leave care at 18 to be extended into their twenties, including contributions to expenses and accommodation.

Waterhouse highlighted the serious results of children not having a voice in care and of children and young people's voices being ignored or sidelined in schools or the community. This evidence contributed to persuading politicians in Wales (unlike those in England) that children and young people should not be dealt with as recipients of services but instead were to be heard, and could have expectations about the delivery of a system for them. In Sen's terms, we could say they were to have *agency*.

IV.1 A Different and Principled Youth Policy for Wales

The Expert Group on what such a system should be was chaired by a prominent voluntary youth project worker from the Rhondda Valley, Margaret Jervis. (She exemplifies commitment: her project had begun small, with one wage paying two people to work round the clock to get disaffected young people involved.) The group report was launched by the Education and Lifelong Learning Minister, Jane Davidson, and the Health Minister, Jane Hutt, in 2001, and adopted unanimously by the Welsh Assembly.

The politicians did not decide the title *Extending Entitlement*. One of the group, Howard Williamson, an academic, is agreed to have coined the term, but he admits that he has never read Sen, or other learned authors on the entitlement approach and capabilities. However, Williamson, like some other members of the group, did contribute on youth issues within UK and European working groups, for whom the relational rather than the distributional notion of poverty was being taken forward by operationalizing social exclusion as the framework for policy. It is no accident that Sen's concepts and their interactive development are widely referred to within the social exclusion literature (Gaudier 1993; also Gore 1993; Rodgers *et al.* 1995: 8–9, ch. 3). I think it is fair to say that the overlap between those interested in youth services, and academics researching poverty and deprivation, was not very great, but there is an obvious coincidence of interests, for example in health and education, that promoted cross-fertilization through debate rather than published work.

More interesting than the Welsh group's reading matter was the explicit principled stance they adopted. Clearly contrasting the Welsh approach from that taken by the UK government for England, Williamson argued in his weekly magazine diary for December 2003 that:

> Deliberations in Wales started at a different point—not the delinquents, druggies and drop-outs . . . but "sorted" young adults who are well equipped for participation in the labour market and civil society. We considered their past experience. It had comprised at its

heart a decent education but it was likely to have been much more than this. There had also been strong family support, away-from-home experiences, access to sport and leisure opportunities, contact with other cultures, and the availability of information services and new technologies. We maintained that young people in difficulty and causing difficulty had had little of this; it should be no surprise that they were parochial, introspective, often alienated and sometimes racist and homophobic: both troubled and troublesome. The public policy challenge, therefore, was to extend a broad "entitlement" to this range of opportunities and experiences to those who were unlikely to access them by "private" or other means. (Williamson 2007)

Within *Extending Entitlement* itself, the contrast with UK government policy for England is similarly explicit:

But government policies have tended to focus on only one manifestation—the offender, the homeless young person, the school refuser and so on, and that particular policy context defines the problem rather than listening to the young person to see things more in the round and address the underlying causes. Effective responses depend critically on understanding the different priorities which young people may attach to problems in their lives. Thus a drug user may well not see drug dependency as the central problem—the problem is destitution, unemployment, family rejection or lack of future prospects—for them drug or alcohol use is the solution or means of escape. So without attending to these issues intervention around drug use is unlikely to help. Changing this compartmentalised response to young people is not easy to achieve[.] (NAW 2000: §§2.32–2.33)

In addition, the report argues for society's responsibility to young people (rather than the young person's responsibility as a recipient of targeted support) and against stigmatized support to individuals:

[W]here this support [by home, family, friends and school] is lacking society has a responsibility to fill the gap—if it does not it will both fail the young person and lead to much higher costs on public services later.... Young people who need extra support should not be stigmatised—many young people from stable families encounter problems and all young people need access to challenges and opportunities beyond their home and community.

(NAW 2000: §1.5)

The differences between England and Wales—at both the theoretical and policy level—are probably overstated. Indeed, policy reports have been important binationally (e.g. SEU 1999, 2000). England has already imitated Wales by appointing a Children's Commissioner, and more convergence is likely (for example, policy on child poverty: see Balls 2007). *Extending Entitlement* is also well known and admired in Europe, and sits within Council of Europe guidelines. This could be because the Welsh Minister, Jane Davidson, has for several years been the UK's political representative in Europe on youth issues, and advocates youth policy from a distinctly Welsh position. It could, however, be argued that the context of Wales favors different implementation: as we shall see, non-devolved issues may have some better outcomes for Wales, which supports this possibility.

IV.2 Form versus Reality

It is useful to distinguish form and reality in any development policy. In particular, the structure of twenty-first century policy and its funding and implementation through partnerships of voluntary and statutory agencies can result in forests of documentation (strategy, plans, service-level agreements, bids for funding, targets and so on) and a related ongoing information problem. This problem, also known as mulletry, involves ceaseless collection and collation of data and information for measuring professional activity, competence or outcomes, and for baselines (to assist future evaluation, quality assurance or performance measures). A tendency for micro-management, fueled by internal markets and a managerial model based on mistrust, has exacerbated these unintended consequences bi-nationally. A virtual world can be created with little real benefit, sometimes perverse incentives, and serious barriers to functionally illiterate or form-averse citizens. Drakeford (2007: 177) argues that a high-trust cooperative approach in Wales protects against this, though I am not convinced.

The excesses of over-documentation, especially from restructuring, were initially damped in *Extending Entitlement* by the argument that "new structures are not needed—rather a local network of quality services guided by a clear vision of how young people's needs will be met" (NAW 2000). This argument for cementing quality was overlooked, however, and new forms were introduced, with the perverse effect of increasing access for socially included young people rather than the excluded (Haines *et al.* 2004).

The avoidance of such unintended consequences, and the sustaining of a focus on supporting real young people, will depend in part on practices in partnerships. Early evaluation of partnerships with a central role in delivering *Extending Entitlement* seems to bode well. Basically, this central role in each of the 22 local authorities falls to a Strategic Unit and a Young People's Partnership. Haines *et al.* (2004) found that, in general, the Young People's Partnerships have found the document and later guidelines very clear, and the results positive.

IV.3 Grounds for Guarded Optimism

I asked some politicians and advisers for specific examples of their principled strategy in action for children and young people, and they highlighted the incorporation of young people into civil and school government (providing the delivery of a system, and making sure that their views are heard and acted on), free swimming for all children up to 16 years in the school holidays, and the free breakfast scheme.

The last of these is held up as a good example of progressive universalism and collective commitment (see WAG 2004). It is voluntary in that the primary schools

have to sign up for it—about half have done so—but it is universal in its delivery. Breakfasts are free to all children, which sends out a powerful anti-stigma message. The social benefits are seen as giving all children, without stigma, a good breakfast (some were arriving unfed and early), and new opportunities for socializing and play.

The new forms of incorporation in schools and civil society can usefully be understood as relating to *agency* (a theme in Sen's work which will be taken up in the next section). The Welsh school system remains robustly state comprehensive, without performance rankings in league tables, which Wales scrapped, and with less testing. The rate of school exclusion is lower in Wales—and much lower in five of the six Valleys authorities—than in England, though it is higher than in Scotland (Kenway *et al.* 2005).

Young people's agency in schools is mainly developed through systematic inclusion in youth councils, whose budget and views must be considered by school governors. There are statutory powers for students to have transitional support from primary into secondary education. Young people are included in the Schools Inspectorate as Peer Inspectors. To carry out these functions young people receive skills training in putting forward their views, and this has spin-off benefits, especially if—and it is a big *if*— the most disadvantaged, troubled or troublesome are recruited into the structure.

The agency of young people within civil society is taken forward through local area youth forums which feed into the local authority forum, and there is also a Children and Young People's Assembly for Wales online (Funky Dragon n.d.). One aim is that young people will allocate a proportion of the youth services budget. The Children's Commissioner has been actively involved, for example in the evolution of an advocacy service (Pithouse 2006; WAG 2007). These examples also implicitly relate to the protective provision of a social wage for the benefit of young people, and to social capital—a measure of collective responsibility to ensure the development of the whole child.

There are some grounds for optimism about the long-term impact of the framework and operation of the above policies, although evaluations of their reach to the disaffected or status-zero children and young people are, at best, only tentative (Haines *et al.* 2004: 23, 101; Lynch and Murphy 2007). The framework for developing the whole child has been put in place, and undoubtedly benefits from the abiding strength of the statutory sector, especially education, in Wales.

This section has traced the emergence of the seminal youth policy framework *Extending Entitlement* and shown clearly that the term *entitlement* bears little direct relation to Sen's definition of an individual's command over goods and services. However, *Extending Entitlement* does address young people's *realized functioning* and *capabilities*—what young people can do and be—and can be interpreted within Sen's theorization (Sen 1992, 1999), beginning with its solid description of the complex issues around troubled and troublesome youth.

In the following sections, I shall look further at reality; first, through the adaptation of a bi-national policy which has particularly affected deprived areas, and then through adult–youth relations in public space, of which both young people and poor people tend to be high users. Some advantages in the Wales policy context emerge, but also a gap in understanding social relations which undermines young people's capabilities.

V. Young People in South Wales: Adaptation of Bi-National Policy

Progressive universalism, as practiced by the Welsh Assembly, intends that people should be beneficiaries on the basis of living in Wales, and not on the basis of good behavior. Furthermore, the ethos of youth policy is encouragement and collective responsibility (NAW 2000: §§10.3–10.11). We might expect, then, that the treatment of troublesome or criminal behavior of young people might differ in Wales from that in England, where much of the criminal justice rhetoric has been punitive and exclusionary in character. This applies also to anti-social behavior, which has become a UK government concern, and has prompted legislation that is particularly judgemental towards youth. However, policy in this area is bi-national, and the Welsh government can influence it only indirectly. Despite this, it has been suggested that Wales has more positive outcomes with respect to youth offending.

In section V.1, I shall ask to what extent the devolved Welsh social policy context can have an impact on the results of a bi-national policy; in section V.2 I shall examine some empirical research about the social relations of adults and youth in public places when adults complain about young people's behavior.

V.1 Area Deprivation, Anti-Social Behavior and Control

The argument that something needed to be done about non-criminal but disruptive behavior emerged in part from debates about regeneration (e.g. in Hills 1998). In fact, it is possible to draw up a typology of regeneration areas, reflecting the lead agency's (or their funder's) predominant interests and perception of the primary cause of social exclusion. One such typology distinguishes between a *control- or crime-reduction-led* model, a *welfare-led* model, and a *community-enterprise-led* model of regeneration (Kynch *et al.* 1999). The leading player in control-led

regeneration was the Home Office, which oversaw a program in 20 cities and boroughs between 1988 and 1995. The central objectives were the reduction of crime and the fear of crime. In this model the creation of community safety is seen as the necessary prerequisite to the re-establishment of a sense of community, and to other forms of development.

Control-led regeneration practices drew on the notion that taking minor nuisance seriously will prevent more serious crime (Wilson and Kelling's 1982 *broken windows* argument). Tactics included surveillance by residents or CCTV, zero-tolerance policing, and the targeting and disruption of troublesome groups. Moreover, a tendency to focus upon perceived threats (in addition to crime itself), such as groups of young people hanging about the streets or aggressive or noisy neighbors, was argued to be the community's expressed priorities, and this was cited as the justification for the UK government focus on behavior. Indeed, this *appeal to community* (Crawford 1997) could become an excuse for seeing control as an end in itself.

Anti-social behavior legislation was introduced through the Crime and Disorder Act 1998 and Anti-Social Behaviour Act 2003, among others. Importantly, the former Act introduced Crime and Disorder Reduction Partnerships to all local authority areas and an obligation on authorities to consider the impact on crime and fear of crime across policy, introducing leverage even into the welfare- and enterprise-led regeneration areas. In Wales, it was argued that safety was a universal issue and addressed social insecurity, and the partnerships were renamed Community Safety Partnerships. This relabeling also reflected a resistance to the growing tendency noted by criminologists for social problems to be reconfigured into crime problems, or the criminalization of social policy (see Simon 2007).

V.2 Is Wales more Cooperative than Coercive?

The former chair of the Youth Justice Board, Rod Morgan, observes in a lecture (Morgan 2007) that there have been two worrying trends in recent years. First, over 15 years, the number of children and young people being held in custody in England and Wales has doubled (despite an existing target to reduce the number in custody), and secondly, although crime has been falling, over the last three years there has been an increase of about 25% in the criminalization of young people (see also Allen 2006). However, Morgan holds up the treatment of young offenders in South Wales as an example that has bucked this trend, and contrasts South Wales with West Yorkshire, a similar post-industrial part of North England.

Morgan notes similarities between the areas as formerly heavily industrial mining areas with "significant social problems; poverty, unemployment. But the criminalization trend is very different in those two areas." West Yorkshire has increased

the number of children and young people criminalized by 61% in three years; South Wales has reduced it by 3%. It is possible, he suggests, that South Wales, and the Chief Constable, take a different strategic line in partnerships on estates. He observes that Chief Constable Barbara Wilding "has made various public statements saying that she regards the use of certain tools in the antisocial behaviour toolbox as very much a last resort and an indication of failure if they have to be resorted to. . . . [S]he doesn't think that arresting young people, for certain sorts of behaviour, is necessarily the most constructive way of dealing with antisocial behaviour." Wilding has also worked closely with the Children's Commissioner, Expert Group member Keith Towler, and has involved young people in designing diversionary activities for youth at risk (South Wales Police 2006).

More evidence that Wales is acting a little tangentially to the bi-national line can be found in the UK government's curt response to the Welsh Affairs Committee's endorsement of Wilding's approach. The Committee wrote: "We agree that ASBOs [anti-social behavior orders] should be issued as a last resort and welcome the Welsh forces' proportionate and appropriate approach to anti-social behaviour." To which the government retorted, "We do not agree that ASBOs should be used as a last resort" (House of Commons Welsh Affairs Committee 2005 App. 1, §15). (The ASBO is a civil order based on levels of evidence lower than in a criminal case; a breach of an ASBO is an imprisonable offence, although the behavior in question is not a crime.)

What South Wales was doing was to negotiate creatively, using warnings or contracts, with young people and parents, in schools or the community, or providing activities to gain young people's cooperation in changing behavior. Importantly, the Community Safety Partnerships often work closely with Children and Young People's Partnerships, the latter being central for delivery within the *Extending Entitlement* ethos.

Whether this can be sustained is another matter, for the rhetoric and media portrayal of young people as very troublesome also crosses national boundaries.

VI. Youth Annoyance: Social Relations in Public Spaces

Although I have argued that devolution may work positively for young people in Wales, an important aspect of being a young person may have been overlooked, and that is the dynamics of incidents involving adult and young people in shared public space. As Sen has argued (and as we argued in Kynch and Sen 1983), social relations are often characterized by both cooperation and conflict, and bargaining

power over who gets or does what is affected by a person's fallback if the relationship is absent or breaks down (e.g. a young person is ejected from home by parents, or not tolerated in public space). Sen's approach also emphasizes how the perception of needs and perceptions of contributions affect bargaining outcomes. Agarwal (1994: 54–60) draws out aspects of bargaining, such as the possibility and forms of resistance used by the powerless to discomfort the powerful, which are relevant here.

This section will draw on research about young people's behavior in public spaces in order to investigate adult–youth relations as cooperative conflicts. Although little attention has been paid to relations in the unstructured setting of public space as opposed to the household, public space is a good place to discover how adults and young people rub along. Indeed, public space provides social and outdoor experiences in line with a recommended entitlement—or realized functioning—in the policy framework of *Extending Entitlement*. However, this view might be thought a little controversial, because the Home Office's British Crime Survey regards the presence of teenagers hanging out as an indicator of disorder, and a third of respondents in South Wales thought it problematic (Aitchison and Kynch 2004).

The concept of youth is emotive and young people have always been seen as rebellious and challenging. One issue is whether they are over- or under-controlled, over- or under-protected (Loader 1996). The media play a powerful part in constructing and sustaining negative images of children and young people as engaged in disorder and types of behavior which threaten society—a continuous *moral panic* (Cohen 1972; McRobbie 1994: 199). Social constructions of youth tend to divert attention away from, or misconstrue, the mundane nature of their regular activities in public spaces.

VI.1 Youth Annoyance in Swansea: Some Research Findings

In 2002, I led a collaborative team of researchers from the Welsh Assembly, a Community Safety Partnership and Children and Young People's Strategy Unit in researching the perceptions and experiences of *youth annoyance* in Swansea, a non-metropolitan city of South Wales (Kynch *et al.* 2004). Swansea has a strong reputation for positive work with young people.

"Youth annoyance" was a local police and council term for the activities of young people in public locations and spaces which cause offense to others. It encompasses crime, disorder, anti-social behavior, incivilities and simply "hanging around". An incident of youth annoyance involves an adult–youth relationship in which age is the defining social symbol of power, and adult perceptions are given credence.

The Research Samples

We looked at three different viewpoints: young people (young people interviewed face to face: 113 on the street and 77 in schools), expert practitioners (31 interviews) and adult residents (924 postal responses from a representative panel).

The *street sample* was a broad cross-section of young people who spend time on the streets and are likely to become involved in intergenerational incidents. They were sought out and approached where they hung out (with help from detached youth workers or Youth Offending Team (YOT) staff), or in drop-in facilities. We went out on the streets in the wind, the wet and the cold, and no doubt interviewees huddling in these weathers included vulnerable young people who felt unable to be at home, in both affluent and deprived areas. We also observed at first hand the excellent relationships of detached youth workers with young people in bleak and deprived areas.

All young people were likely to be high users of public space, with nine out of ten saying they had hung out on the streets in the previous month. Many were out on most days, for several hours a day and at all hours of the day and night, and this constant presence attracted frequent complaints. Public space, and how space is shared, mattered to them, as drop-in center staff recognized. The street sample was usually socially engaged, more so, if anything, than the school sample, being marginally more likely to play in teams or have structured activities. More of them also had some paid employment.

Types of Youth Annoyance: Criminal, Anti-Social, Disputed and Negotiable

As I noted in section V, legislation on anti-social behavior rests on what are said to be communities' shared perceptions and priorities. However, our research suggests that this gives too much weight to the minority of adults who are offended or upset by young people's mere presence, especially in groups.

What do we classify as anti-social behavior? The opinions of adults, experts and young people are summarized in Figure 6.1. We found a strong consensus on what was criminal, and on what was not criminal but justified a complaint. However, there was considerable disagreement about two other types: *disputed behavior*, perceived as annoying or troublesome depending on the situation or context, or property rights, and *negotiable behavior*, activities which were usually tolerable or sometimes even pro-social.

With regard to the examples of disputed and negotiable types of behavior, a substantial minority of adults thought them annoying, but nearly all the young people said they were all right and did not justify a complaint. The potential for conflicts was high, and we shall look later at how complaints worked out.

Criminal activities: e.g. mugging and attacking people, using drugs in public, stealing, operating in gangs and vandalism.
"Anti-social" behavior: e.g. throwing things, loud swearing or abusive language, graffiti (tagging) and drinking alcohol publicly.
"Disputed" behavior (or "threshold" behavior): e.g. being rowdy or boisterous, or kicking footballs against walls or garage doors.
"Negotiable" behavior: e.g. playing loud music, skateboarding in the street, hanging around street corners or parks and playing in the street.

Fig. 6.1. A typology of youth annoyance, based on replies from adults, experts and street young people

Note: Adults and experts were asked, for each of 16 types of behavior, if they would regard them as youth annoyance. Young people were asked if the activities were "OK" or "Not OK and justified a complaint".

For example, skateboarding was popular in Swansea as a street activity, and figured frequently in complaints. Yet it is intrinsically quite healthy, skillful and sociable. The key difference between skateboarding incidents that were negotiated successfully, those that included resistance, and those that left the young person aggrieved, was the adult–youth relations—not the behavior.

Perceptions of Young People's Behavior

The perception of a person as having low value can undermine their bargaining in social relations. As already noted, teenagers are not greatly valued in public space; in fact they are officially labeled a *disorder*. The young people in our study were aware that adults perceived them as a problem, and the street sample tended to overestimate greatly how much of a problem they were perceived to be.

Some of the experts gave examples of misconstrued behavior, and a few showed awareness that young people do like to wind up authority. By establishing the facts in place of misconstrued adult opinion, or maintaining a little humor in the face of provocation, the police, YOT and Youth Service workers could make themselves available as arbitrators when incidents blew up.

Complaints against Young People

Asked about their experiences during the most recent complaint against them, the young people brought out the different ways in which they resolved, or failed to resolve, an incident. The great majority of complaints involved disputed or negotiable behavior, and such incidents are the main concern here.

The young people had a strong sense of justice and showed awareness of how their age was used to manipulate them into relative powerlessness in the community. They also used active and passive resistance in order to persist with their activities and presence in public space.

Cooperation: Compliance and Justice

A substantial group of around half the young people thought there had been a good reason for the adult to complain, and 41% claimed that they had complied with what was asked, occasionally apologizing. They appreciated explanation.

Fair complaints included when neighbors of footballers in the street were "scared in case there was a car crash, or we get run over". A skateboarder said that when skating in the University, "security guards complained. Yes, fair—people trying to work. [We were] moved on—[but it] should be OK at weekends."

Conflict: Resistance and Persistence

Resistance was common, and parents occasionally had to back up the young people when an adult was unreasonable, or called the police. Access to an advocate like a parent was an important part of the young person's fallback. There was little evidence in the narratives that young people intended to stop their activities permanently. Instead they looked for another location or time—for example, skateboarders networked information about where to skate without contestation.

Resistance could deflect conflict when adults seemed to be deliberately targeting young people. The young person might hold their ground, like another skateboarder who told us he was "on a cycle path and had a board set up at the side—a cyclist deliberately knocked it down—obvious what he meant. Didn't even slow down. Attitude really. Not very fair." Passive resistance was also employed, for example a 16-year-old was "standing there; asked by people driving past 'Have you smoked a spliff . . . this and that'. I stayed standing. Completely unfair—they drove away and said sorry to bother me."

However, conflicts could leave the young person expressing aggrieved powerlessness. This is another skateboarder: "Just skateboarding—not my own street—on a road near my house. This woman aged about 40 complained. Not very fair: we weren't really doing much—we left. Next time we went there she rang the police without coming out. The police booked us for skateboarding on the road. They took our names, and wrote to our parents—saying 'they could have been doing drugs and alcohol'—but we were just skateboarding."

Breakdown and Adult Force

Many adults may be unaware of how young people see them behaving and their role in escalation and breakdown—for example, a hanger-out on a pavement found a complaint "Not very fair—shopkeeper told me to fuck off". Rudeness and adult force permeated breakdown as everyday experiences for some: "This old man came out: 'stop kicking the can'. I shouted back to 'shut up'." As one 11-year old put it, "adults shouldn't go round to a person and give them the violence. [We] shout violence back at adults. Adults start it."

In summary, young people in Swansea, as elsewhere in the UK, often annoy adults. Conflicts in adult–youth relations in public spaces reflect adult perceptions and low adult valuations of young people outside the home, and often the young people exhibit forms of overt and covert resistance to the apparently arbitrary exercise of power. It should not be forgotten that these young people will be tomorrow's adults, and inclusion of a capability to be present and active or sociable in public space is encouraged by positive relations there, and the collective responsibility for bringing these about.

VII. Conclusions

I have argued that the history and context of post-industrial Wales has been conducive to devolved policies dependent on collective responsibility for the well-being of its residents, on a universal basis, with extra support for the most disadvantaged. Furthermore, devolution in Wales has arguably encouraged imaginative use of social policy, even though other important policies concerned with the economy and crime and disorder have not been devolved.

Welsh youth policy, as encapsulated in *Extending Entitlement*, caught a moment when the policy-makers made their expectations clear, and youth researchers communicated in a meaningful way as a contribution to a policy agenda. This policy can be interpreted in terms of realized functionings of young people and capabilities, with an emphasis on cooperation and encouragement. In particular, the Welsh Assembly has shown a commitment to deliver a system which effectively addresses the agency of young people.

Shortcomings may lie in the availability of resources, in form versus reality, and in under-theorization. Most importantly, there are doubts about the extent of support to disadvantaged young people, although socially included youth have gained.

Regeneration of deprived post-industrial areas has been affected by a bi-national policy on crime and disorder, and the Wales policy context appears to have resulted in more positive outcomes for young people in terms of criminalization. I have suggested that this could be because partnerships under bi-national policy and devolved Welsh policy work closely together and are affected by the ethos of *Extending Entitlement*.

Although the Welsh policy looks promising, its conceptualization of adult–youth relations is weak, and is prey to the rhetoric of moral panic. I examined incidents of adults becoming annoyed by young people in public space as social relations characterized by both cooperation and conflict. This conceptualization facilitates

attention to both adult and young people's bargaining and weaknesses when relations break down, and could help policy to deliver more substantial freedom for young people to achieve their potential.

This reinterpretation of policy and support for young people in South Wales has been an opportunity to demonstrate the breadth of Amartya Sen's contribution to understanding development, and his indirect beneficial influence on social policy and its wider potential.

References

Agarwal, B. (1994), *A Field of One's Own: Gender and Land Rights in South Asia* (Cambridge: Cambridge University Press).

Aitchison, A., and Kynch, J. (2004), *Crime and Victimisation in Wales: Results from the British Crime Survey 2001/02*, available at <http://new.wales.gov.uk/dsjr/research/bcscrimeandvictimisation/report?lang=en>.

Allen, R. (2006), *From Punishment to Problem Solving: A New Approach to Children in Trouble*, available at <www.kcl.ac.uk/depsta/rel/ccjs/pubs.html>.

Balls, E. (2007), "Childcare and Child Poverty–Delivering Solutions", presentation to Daycare Trust Conference, London, 13 June; available at <www.daycaretrust.org.uk/mod/fileman/files/Ed_Balls_Speech.doc>.

Beddoe, D. (2004), "Women and Politics in Twentieth Century Wales", Welsh Political Archive Annual Lecture; available at <www.llgc.org.uk/index.php?id=505>.

Cohen, S. (1972), *Folk Devils and Moral Panics* (London: Penguin Books).

Cooke, B., and Kothari, U. (eds) (2001), *Participation: The New Tyranny?* (London: Zed Books).

Crawford, A. (1997), *The Local Governance of Crime: Appeals to Community and Partnerships* (Oxford: Clarendon Press).

CWM (n.d.) *South Wales Coalfield Further Reading List*, available at <www.agor.org.uk/cwm/further_reading.asp>.

DOH (2000), *Lost in Care*, Department of Health, Series Number HC Session 1999–00 201, available at <www.dh.gov.uk/enPublicationsandstatistics/PublicationsPolicyAndGuidance/DH_4003097>.

Drakeford, M. (2007), "Social Justice in a Devolved Wales", *Benefits: The Journal of Poverty and Social Justice*, 15: 171–8.

Funky Dragon (n.d), *Children and Young People's Assembly for Wales,* available at <www.funkydragon.org>.

Gaudier, M. (1993), *Poverty, Inequality, Exclusion: New Approaches to Theory and Practice*, Série bibliographique no 17 (Geneva: Institut international d'études sociales).

Haines, K., Case, S., Isles, E., Rees, I., and Hancock, A. (2004), *Extending Entitlement: Making it Real* (Cardiff: Welsh Assembly Government).

Hills, J. (1998), *Persistent Poverty and Lifetime Inequality: The Evidence*, CASE report 5 and HM Treasury Occasional Paper 10 (London: ESRC Centre for Analysis of Social Exclusion, London School of Economics).

House of Commons Welsh Affairs Committee (2005), *Police Service, Crime and Anti-Social Behaviour in Wales*, available at <www.parliament.the-stationery-office.co.uk/pa/cm200506/cmselect/cmwelaf/514/51404.htm>.

Istance, D., and Williamson, H. (1996), *16 and 17 Year Olds in South and Mid Glamorgan Not in Education, Training or Employment (Status Zero)* (Swansea: Mid Glamorgan Training and Enterprise Council).

Kenway, P., Parsons, N., Carr, J., and Palmer, G. (2005), *Monitoring Poverty and Social Exclusion in Wales 2005 with update to 2007*, available at <www.poverty.org.uk/summary/reports.shtml#wales>.

Kynch, J., Drakeford, M., and Hinshelwood, E. (2000), "The Excluded of the Excluded: A Pilot Study of Processes of Individual Exclusion within Area Regeneration", unpubl. report, CDS (University of Wales, Swansea).

—— Maguire, M., Hinshelwood, E., and Vanstone, M. (2000), *Community Involvement in Gurnos and Galon Uchaf*, SOCSCI Working Paper 7 (Cardiff: SOCSCI).

—— Portwood, J., and Timmins, S. (2004), "Youth Annoyance in Swansea", unpubl. summary of joint research by CSU, the Welsh Assembly, the Safer Swansea Partnership, and the Children and Young People Strategy Unit, the City and County of Swansea.

—— and Sen, A. (1983), "Indian Women: Well-Being and Survival", *Cambridge Journal of Economics*, 7: 363–80.

Lanjouw, P., and Stern, N. H. (eds) (1998), *Economic Development in Palanpur over Five Decades*. (Oxford and New Delhi: Oxford University Press).

Loader, I. (1996), *Youth, Policing and Democracy* (Basingstoke: Macmillan).

Lynch, R., and Murphy, S. (2007), *A Process Evaluation of the Welsh Assembly Government's Primary School Free Breakfast Initiative* (Cardiff: Institute of Society, Health and Ethics, Cardiff University).

McRobbie, A. (1992), *Postmodernism and Popular Culture* (London: Routledge).

Morgan, K. O. (1995), *Modern Wales: Politics, Places and People* (Cardiff: University of Wales Press).

Morgan, Rhodri (2006), "A Welsh Recipe for 21st Century Socialism", available at <www.compassonline.org.uk/article.asp?n=338>.

Morgan, Rod (2007), *Rebalancing Youth Justice*, 2006/7 New Developments in Criminal Justice Lecture Seminar Series; available at <www.kcl.ac.uk/depsta/rel/ccjs/pubs.html>.

NAW (2000), *Extending Entitlement: Supporting Young People in Wales*, available at <http://new.wales.gov.uk/topics/educationandskills/policy_strategy_and_planning/extending_entitlement/eepublications/supportyoungpeople?lang=en>.

Pal S., and Kynch, J. (2000), "Determinants of Occupational Change and Mobility in Rural India", *Applied Economics*, 32: 1559–73.

Pithouse A. (2006), "A Common Assessment for Children in Need? Mixed Messages from a Pilot Study in Wales", *Child Care in Practice*, 12(3): 199–217.

Rodgers, G., Gore, C., and Figueiredo, J. B. (eds) (1995), *Social Exclusion: Rhetoric, Reality, Responses* (Geneva: ILO and UNDP).

Room, G. (ed.) (1995), *Beyond the Threshold: The Measurement and Analysis of Social Exclusion* (Bristol: Policy Press).

Sen, A. (1992), *Inequality Re-Examined* (Oxford: Clarendon Press).

—— (1999), *Development as Freedom* (Oxford: Oxford University Press).

SEU (1999), *Bridging the Gap* (London: Social Exclusion Unit, Cabinet Office).

SEU (2000), *National Strategy for Neighbourhood Renewal, Policy Action Team 12—Young People* (London: Social Exclusion Unit, Cabinet Office).

Simon, J. (2007), *Governing through Crime* (Oxford: Oxford University Press).

South Wales Police. (2006), "Providing Young People with Someone to Listen and Something to Do", available at <www.south-wales.police.uk/fe_news/news_details.asp?newsid=2209>.

WAG. (2004), *Primary Schools Free Breakfasts Initiative Paper*, available at <http://new.wales.gov.uk/topics/educationandskills.policy_strategy_and_planning/schools/breakfast_initiative/?lang=en>.

—— (2007), *New Service Model for Delivering Advocacy Services for Children and Young People*, available at <http://new.wales.gov.uk/topics/childrenyoungpeople/childrenfirst/?lang=en>.

Williams, G. A. (1985), *When Was Wales? The History, People and Culture of an Ancient Country* (London: Penguin Books).

Williamson, H. (1997), "Status Zero Youth and the Underclass", in R. MacDonald (ed.), *Youth, the "Underclass" and Social Exclusion* (London: Routledge).

—— (2007), *The Thoughts of Chairman How* (Leicester: National Youth Agency), available at <www.nya.org.uk>.

Wilson, J. Q., and Kelling, G. L. (1982), "Police and Neighborhood Safety: Broken Windows", *Atlantic Monthly*, March: 29–38.

CHAPTER 7

COUNTRY PATTERNS OF BEHAVIOR ON BROADER DIMENSIONS OF HUMAN DEVELOPMENT

GUSTAV RANIS
FRANCES STEWART
EMMA SAMMAN

I. Introduction

Human development (HD) has been defined as "a process of enlarging people's choices". Although often equated with the Human Development Index (HDI), as a combined set of measures of education, health and incomes, the HDI represents a basic and reductionist version of HD. As Amartya Sen has pointed out,

human development encompasses much more than is included in the HDI. As he puts it:

> (I)t would be a great mistake to concentrate too much on the Human Development Index or on any such aggregative index . . . These are useful indicators in rough and ready work, but the real merit of the human development approach lies in the plural attention it brings to bear on developmental evaluation, not in the aggregative measures it presents as an aid to the digestion of diverse statistics. (Sen 2000: 22)

In earlier work (Ranis *et al.* 2006), we extended the measurement of HD to 12 important categories of life and proposed plausible indicators within each category. We then eliminated the indicators highly correlated with others in the same category, leaving us with 40 relatively independent ones. We subsequently eliminated those highly correlated with the core HDI and were still left with 32. This suggests that a full assessment of human development requires us to move beyond the HDI. Moreover, this already implies that performance across different dimensions of human development need not be consistent, and that country patterns of performance may differ. Some countries may do well on some dimensions and poorly on others; others may do well or poorly on all dimensions.

The central aim of this chapter is to adopt a more expansive definition of HD than that of the HDI, in order to explore such alternative patterns of country behavior. We are interested in identifying countries which, for one reason or another, seem to do particularly well on one dimension and less well on others, or particularly badly on one dimension and better on others, as well as managing to do well on all, or failing to do well on any.

Countries may show different patterns of performance because—with limited resources and capacities—they *choose* to emphasize one dimension rather than another (e.g. they choose to promote economic growth at the expense of social ties, or political freedoms); or because they face *constraints* which prevent success on one dimension but allow success elsewhere; or because particular patterns have resulted from their *history* or *culture* (e.g. a culture which involves strong social ties, or one, like Costa Rica, where the basic HDI elements have long been promoted). There are also causal connections across some dimensions (notably, for example, between basic human development and economic aspects) that constrain the range of possible behavior over time, and also help determine the pattern of choices observed.[1] This chapter seeks to identify actual patterns of country behavior so that one can begin to consider how far the outcomes are a matter of choice, of constraints, or of history and culture.

Our first task is to identify the dimensions of HD to be investigated. This is the subject of section II. We subsequently present the methodology adopted for classifying country behavior according to the dimensions chosen (section III).

[1] The connections between HDI and economic growth are explored in Ranis *et al.* (2000) and Boozer *et al.* (2003). Much work has also been done in exploring connections between political freedoms and economic growth, and political freedoms and HDI.

Section IV presents the main results, while section V concludes by discussing some implications of the results.

II. Dimensions of Human Development

Identification of a unique and correct list of all the possible dimensions of HD is an impossible task. As is well known, Sen himself has always refused to identify an exhaustive list of freedoms or capabilities, i.e. of those beings and doings that people have reason to value. However, many philosophers, from the ancient Greeks onwards, have attempted to provide an answer to the question of what constitutes the good or full life, and have come up with numerous responses. Alkire (2002), for example, summarizes 39 attempts over the years 1938–2000 to specify the characteristics of a full life, based on a variety of philosophical justifications. Drawing on six recent approaches,[2] we found that a number of common categories could be identified: bodily well-being, material well-being, mental development, work, security, social relations, spiritual well-being, empowerment, political freedom, and respect for other species, the last appearing only in Nussbaum (2000). However, from our perspective, this list is excessive for two reasons: first, for some aspects (notably spiritual well-being and respect for other species) data are not available; of even greater importance, adopting too many categories would make it difficult to classify countries' behavior sensibly.

Hence we decided to reduce our set of categories to four. Our objective in choosing these four was that, first, each should represent an important aspect of human choice and one that is broadly independent, at least conceptually if not causally, of the others; and, second, that each should encompass a large proportion of the categories identified by others as constituting essential aspects of full human flourishing. In the light of these two considerations, we chose the following four categories:

1. *Basic HD*. Rather than using the HDI to measure this, we use the under-five mortality rate, partly because we want to exclude income per capita as it appears in the next category, and partly because of the extensive availability of data for the under-five mortality indicator. In fact, the under-five mortality rate is highly correlated with the HDI and with adult literacy (0.8789 and 0.7393 respectively for 113 countries in 2002), so it can be taken as representative of these indicators.[3]

2. *The economic aspects*. In our interpretation, this encompasses income per capita and unemployment to represent economic performance at a point in time,

[2] These are: Rawls (1972); Finnis *et al.* (1987); Doyal and Gough (1993); Narayan-Parker (2000); Nussbaum (2000); and Camfield (2005).

[3] All correlations are based on the Spearman rank-order method.

and growth in per capita income and the GDP cycle to represent performance over a longer period. These aspects therefore broadly summarize an economy's success in providing incomes, employment, growth and economic stability.

3. *Social and community relations*. We include a quite large and disparate set of variables here to represent different facets of success in achieving a flourishing community and good social relations, including measures of income distribution, the perceived importance of family and friends, tolerance of neighbors, and gender empowerment, as well as (negatively) the male suicide rate. The crime rate would have been another good variable, but was precluded by data limitations.

4. *Political freedoms and stability*. This category includes an index of political and civil liberties, a measure of the rule of law, and a measure of collective political violence.

In shorthand, we term these four categories basic HD and the economic, social and political dimensions of HD. Of the ten categories which we identified above as the main dimensions of some of the major philosophical efforts to identify conditions for human flourishing, only spiritual well-being and respect for other species are entirely excluded. In principle, mental development is included in basic HD (and education is highly correlated with it); aspects of work are included in the economic category; aspects of security in the political category; bodily well-being in basic HD; material well-being under economic aspects; social relations under social aspects; and empowerment and political freedom in the political category. An important dimension that is omitted at this stage is respect for the environment.[4] We also acknowledge that there is a good deal of arbitrariness in the selection and assignment of variables. Table 7.1 summarizes the categories and Table 7.2 gives the details of the measures adopted and their sources.

III. Method Adopted for Classifying Countries

We start by acknowledging that an exercise of this kind unavoidably involves many arbitrary decisions. Thus our country classification system should be seen as suggestive rather than conclusive. In future work, it will be important to investigate how far the use of different indicators, aggregation procedures for each category,

[4] For a composite measure of environmental sustainability see Yale Center for Economic Law and Policy and CIESIN (2005). An index to represent this showed only a low correlation with the HDI over 90 developing countries (0.2553), so it is an important independent dimension which ought to be incorporated.

Table 7.1. The four categories of human development

Category	What it consists in	How it is measured[a]
Basic HD	Bodily well-being Mental development	Under-five mortality rate
Economic dimensions	Material development Work	Income per capita (PPP) Growth in per capita GDP (10-year average) GDP cycle (20-year average) Unemployment rate
Social dimensions	Social relations	Income distribution Importance of family Importance of friends Tolerance of neighbors Gender empowerment (GEM) Male suicide rate
Political dimensions	Empowerment and political freedoms	Index of political and civil liberties Index of rule of law Collective political violence

[a] All data are for 2002 unless otherwise indicated. See Table 7.2 for more detail on indicators and data sources.

and classification procedures for each country's performance would alter the results.

A greater problem that may affect the results is that data are missing for a large number of countries. We proceeded with our approach despite large gaps in data availability. We provide some indication of the extent to which these gaps appear to matter, but also point to the need for far better data coverage—particularly in the case of poorer countries—to arrive at a more robust set of conclusions.

Our methodology has four stages: first, we identify the countries of interest; second, we develop a procedure to identify a summary indicator to represent each category on the basis of the several indicators presented above; third, we develop a method to classify countries as high, medium or low for each of our four categories; and finally, we adopt a classificatory system for countries when examining their overall performance on the indicators.

III.1 Choice of Countries and Analytical Categories

Because the aim of the exercise is to capture variance among developing countries, we eliminated countries defined as "high income" by the UN. In addition, we excluded countries with fewer than one million inhabitants (based on UN estimates for 2002). We first categorized countries by region—i.e. Sub-Saharan Africa, Latin

Table 7.2. Summary of indicators and sources

Indicator	Definition	Date	Source
Basic HD			
Child mortality rate	Under 5 years old, per 1,000 live births	2002	UNDP (2004)
Economic Dimension			
Unemployment rate		Most recent av. (1992–2003)	ILO (various years)
Per capita GDP	PPP US$	2002	World Bank (2004)
GDP cycle	Avg. annual deviation from mean	1981–2002	Calculated from World Bank (2004) data
Per capita GDP growth	10 year average GDP growth	1994–2005	World Bank (2006)
Social Dimension			
Income distribution	Gini of income	1990–2000 (most recent av.)	UNDP (2004)
Male suicide rate	Per 100,000	2003 (or most recent av.)	WHO (various years)
FamilyVeryImpt	Share indicating family "very important"	Most recent av.	World Values Survey (various years)
FriendsVeryImpt	Share indicating friends "very important"	Most recent av.	World Values Survey
Neighbor Tolerance	Average response to whether would want to live next to various types of people; lower numbers indicate more tolerance	1999/2001	World Values Survey
Gender empowerment measure (GEM)	Composite of gender inequality in parliament, occupational status and income	2002	UNDP (2004)

Political Dimension			
Combined pol rights/civ liberties indicator	Scale of 1–7 with 1 most free; average of 'political rights' & 'civil liberties' scales	2003	Freedom House (2006)
Rule of law	Extent to which agents have confidence in & abide by rules of society; higher numbers indicate better rule	2002	World Bank (2007)
Collective political violence in 1990s	Reflects levels of violence within country & whether excessive civilian targeting, 0-8 with 8 worst	1990s	Marshall (2002)
Classifications			
Landlocked			United Nations designation (UN-OHRLLS, various years)
Distance from equator	Countries are divided into thirds based on distance from equator (n = 41)		World Bank (1999, 2000)
Population levels	Countries are divided into thirds based on population (n = 127)	2004	UNDP (2006)
Life satisfaction	Countries are divided into thirds based on life satisfaction (0-10 ladder, 10 = most satisfied) (n = 48)	1990s	World Database of Happiness (various years)

America, the Middle East, South-East Asia, South Asia, Central Asia and Eastern Europe—and geographical characteristics, i.e. according to whether they are landlocked, and by their distance from the equator. Secondly, we classified countries according to various economic, political and social characteristics: low income and middle income (using World Bank definitions); conflict and post-conflict economies; oil economies; and transition economies. We subsequently classified countries on the basis of average life satisfaction. The aim was to explore whether different country types behave differently with respect to our four categories, although clearly there is some overlap across the different categorizations.

III.2 Obtaining a Single Indicator for each Category

Here we confront the normal problems of devising multidimensional indices. Simple averaging is not possible for indicators using different scales of measurement. The HDI attempts to solve this problem with the shortfall approach, giving each indicator a rank according to the percentage of a country's shortfall compared to the best performers, with the total range being set by the difference between the low and high performers. This puts all indicators on a comparable scale but averaging remains an arbitrary process, both because there is no particular reason why every indicator should be valued equally, and (somewhat paradoxically) because the three HDI component indicators are not valued equally, as the range may differ from indicator to indicator.

We have therefore adopted a different approach. We classified each country for each indicator relative to the median. The median and other order-based statistics were preferred over parametric measures because the distribution of countries for most indicators deviated sharply from normal. Countries were classified as "medium" (M) if they fell within the interquartile range (IQR) for a particular indicator, as "high" (H) if they were above the IQR (in the top 25% of countries), and as "low" (L) if they were below the IQR (in the bottom 25%).

As noted earlier, one major issue was missing data. If we had omitted all countries with missing data we would have been left with a very small sample indeed. To avoid this, we ignored missing data, unless they were missing on *every* indicator in the category. This means that some countries are classified on the basis of fewer indicators than others. For example, Botswana has full data for the economic, political and basic HD categories, but is very deficient in the social category, with data for income distribution and gender empowerment (GEM) but none on the importance of friends and family, male suicide and tolerance of neighbors. Accordingly, its social categorization rests on just the two indicators available.

Only 21 countries—all high and medium performers—had data on all the social indicators, making it important to consider further the implications of absent data. The paucity of data presents particular problems, as the indicators within

each category were selected in part because of their low correlation with other indicators in that category, so we could not reliably infer a country's performance on indicators with missing data from its performance on other indicators within that category.

To determine the extent to which the missing data affected the results in the social category, we recategorized all of the countries in the absence of each indicator in turn. We found that removing any one indicator, apart from the Gini, changed the overall country designation in the social category for less than 10% of countries.[5] However, removing the Gini, for which far more data are available, changed the social category result in over one third of the cases (34%). This finding reinforces the need to treat these results with caution and, when referring to a country's social performance, to specify what the social category represents.

III.3 Classifying Countries with Respect to each Category

When a country's performance was high on all indicators within a category, it was classified as "high" on that category, and similarly for categories in which performance on all indicators was medium or low. Mixed performance in a particular category was more complex. If a country was classified as a mixture of medium, high and low, it was classified as medium. Mixtures of only high and medium performance were labeled as high, and mixtures of only medium and low as low. While missing data was an acute problem, we only gave a country *no* category when data were missing for *all* indicators in that category.

III.4 Classifying Countries Overall

With each country having been assigned a classification for each category, we followed a similar approach to categorizing a country's overall performance in the economic, social, political and basic human development dimensions.

In 27 cases, there were one or more dimensions on which it was not possible to classify countries because of lack of data. This was true particularly of African countries, with 11 out of 42 lacking data on the social dimension. We still proceeded even when only two or three categories were classified, broadly following the same categorization.

Finally, we determined whether a country fared better (or worse) on one particular dimension compared to the other three—for instance, whether it performed better politically than on the other three dimensions, or was deficient in terms of its basic human development. It is these countries which appear to be emphasizing or neglecting some dimensions relative to others. We identified a country as being

[5] The range was between 6 and 8%, depending on the indicator removed.

Table 7.3. Classifying country performance by category and overall

Categorization of dimensions	Country's overall classification	Characterization
All H	H	High throughout
3 H, 1 M	H	High, deficient in one category
3 H, 1 L	M	High, deficient in one category
2 H, 2 M; 2H, 2L; 2 H, 1 M, 1 L	M	Medium
1H, 3M	M	Medium, superior in one dimension
4 M	M	Medium throughout
3M, 1L	M	Medium, deficient in one dimension
2M, 2L	M	Medium, mixed
1M, 3L	L	Low, superior in one dimension
4 L	L	Low throughout

superior in one dimension if it was high in one dimension and medium in all the others, or medium in one dimension and low in the others. Conversely, a country was classified as *deficient* in a particular dimension if it was low in one and medium in three, or medium in one and high in three. We defined a country as *imbalanced* if it was either superior or deficient in any dimension. Others were considered *balanced*.

Table 7.3 summarizes these classification procedures.

The next section gives the results of this methodology.

IV. Main Results

Out of the entire sample of 130 countries, 66 fell into the medium category, 32 into the low category and 32 into the high category, according to the procedures described above. Almost half (55) of the countries showed imbalance, being categorized as superior or deficient in one dimension, thus confirming the fact that performance across categories is often not consistent. The biggest imbalances were on the political and social aspects. In the case of politics, 12 countries were superior on this element in relation to their overall performance, and five were deficient. We would expect parts of the political aspects to be related to other aspects of performance—notably the extent of collective political violence and the rule of law—but others, in particular political and civil liberties, plausibly have only a loose causal connection with the other elements. In relation to social aspects, eight countries showed superior performance and eight were deficient relative to their overall performance. This variable tries to capture the flourishing (or otherwise) of

the community in which people live, although data problems mean that we have, at best, done so only partially. But to the extent we were able to measure this, it again seems reasonable that, while the social dimension is an important aspect of HD in its own right, it will not necessarily be related to the other categories. Six countries showed superior performance in basic HD and six were deficient, while five countries showed superior performance in the economic category and five were deficient.

IV.1 Geographic Performance

As can be seen from Table 7.4, by far the highest proportion and largest number (25) of low performances, on all categories, were in Sub-Saharan Africa. Outside Sub-Saharan Africa, two countries in the Middle East showed low performance—Algeria and Yemen—as did the Democratic Republic of Korea in South-East Asia, and Haiti in Latin America. The region with the most high performances was Eastern Europe, with 11 or half of the countries, followed by Latin America with six—Chile, Costa Rica, Guatemala, Jamaica, Mexico and Uruguay. There were also two in Sub-Saharan Africa—Ghana and Mauritius—two in the Middle East—Libya and Oman—two in East and South-East Asia—Malaysia and Thailand—and three in Central and South Asia—Bhutan, India and Kazakhstan. Landlocked countries performed below average. Performance worsened as countries got closer to the equator, with 38% of countries furthest from the equator in the high category, and only 14% of the countries nearest the equator in that category.

As noted in the previous section, our methodology classified countries as low or high even if they did better or less well on one, or occasionally two, categories. For countries with data for all four categories, only two countries in the world had a consistently high record on all four—Costa Rica and Trinidad and Tobago. Five countries were consistently medium: Bolivia, Brazil, Nepal, Saudi Arabia and Turkey; and only one country, Sierra Leone, had consistently low performance, though Chad, Democratic Republic of the Congo, Iraq, Somalia and Zimbabwe were also consistently low performers, but were missing data on one or two categories.

Imbalanced performance leading to classification as superior or deficient in some dimension affected nearly 60% of Sub-Saharan African countries and about 40% of Latin American and Eastern European countries, with smaller proportions elsewhere. In terms of the nature of the imbalance, we find that a large number of Sub-Saharan African countries were superior in either the social or political dimensions, much more so than elsewhere—indicating that poor basic HD and economic performance were accompanied by better performance on social or political categories. In Latin America, the imbalance came more from deficient performance on the social and economic sides. Put another way, this indicates that

Table 7.4. Geographic performance

	Sub-Saharan Africa	Latin America	Middle East	East and South-East Asia	South and Central Asia	Eastern Europe	Landlocked	Distance from equator		
								Furthest third	Middle third	Closest third
Overall classification										
High	2	6	2	2	3	11	8	5	5	2
Medium	15	14	11	8	7	11	15	7	7	6
Low	25	1	2	1	0	0	13	1	2	6
Proportion imbalanced (%)	57.1	42.9	28.6	18.2	30.0	40.9	52.8	46.2	35.7	50.0
HD superior	2	2	0	0	0	2	1	1	1	1
HD deficient	5	0	0	0	1	0	1	0	1	1
Social superior	6	0	1	0	1	0	4	1	0	0
Social deficient	1	2	0	1	0	4	2	1	1	1
Economic superior	0	1	1	0	1	0	0	1	0	2
Economic deficient	0	3	0	0	0	1	1	1	1	1
Politics superior	10	0	2	0	0	0	7	0	1	1
Politics deficient	0	1	0	1	0	2	3	1	0	0

in some Latin American countries, basic HD and political performance outpace economic and social aspects, while in Sub-Saharan Africa it is political and social performance that is outpacing economic and basic HD. In Eastern Europe, the imbalance came from a combination of social and political deficiency and basic HD superiority.

IV.2 Performance by Country Type

Table 7.5 presents results according to country type.

To start with, we contrast low- and middle-income countries. As is to be expected, low-income countries have more low overall classifications and fewer high ones than those in the middle-income category. But there are still four high classifications in the low-income category—Bhutan, Ghana, India and Mongolia. Similarly, five middle-income countries were classified as low all around—Algeria, Angola, Cameroon, Congo Republic and Swaziland. A very similar proportion of low- and middle-income countries was imbalanced in performance across categories (46% and 42%, respectively). The low-income countries showed a combination of HD deficiency and social and political superiority, following much the same pattern as Sub-Saharan African countries, which, of course, represent a large proportion of these low-income countries. Middle-income countries showed a very different pattern. A large number were socially or economically deficient, while the six countries that were politically superior just exceeded the four which were politically deficient. It thus appears that social performance is negatively associated with levels of per capita income, but there is no systematic relationship between political performance and income per capita. Doing better than on other categories in the politics dimension is quite common for both low- and middle-income countries, but middle-income countries do worse here almost as often as they do on other categories.

A country's population does not seem to have any systematic impact on performance. The countries in the smallest population category had a lower proportion of high performers but also a lower proportion of low performers than the countries with larger populations. The middle-size category had a substantially higher proportion of low performances than either the largest or the smallest countries.

We subsequently explored the performance of countries which have had particular types of experience. Judging by the countries listed so far that have performed poorly, it is already clear that conflict is associated with overall poor performance. Conflict tends to undermine economic and basic HD performance, is obviously associated with political breakdown, and might also be expected to have negative repercussions in the social category (see Stewart and Fitzgerald 2001). This is confirmed by the figures above. Seven out of 13 conflict countries and five out of

Table 7.5. Performance by country type

	Low-income countries	Middle-income countries	Conflict countries	Post-conflict countries	Oil producers	Transition countries	Population Size		
							Top third	Middle third	Bottom third
Overall classification									
High	4	25	0	1	2	6	7	7	15
Medium	19	46	6	1	7	13	26	19	21
Low	27	5	7	5	4	0	9	16	7
Proportion imbalanced (%)	46.0	42.1	30.8	71.4	46.1	21.0	26.2	52.4	46.5
HD superior	2	4	1	0	0	1	2	2	2
HD deficient	6	0	0	1	0	0	1	3	2
Social superior	6	3	2	1	2	1	3	3	3
Social deficient	0	7	0	0	0	0	2	3	2
Economic superior	2	3	0	2	2	0	2	2	1
Economic deficient	0	5	0	0	1	0	1	0	4
Politics superior	6	6	1	0	0	0	0	6	5
Politics deficient	1	4	0	1	1	2	0	3	1

seven post-conflict countries are in the low category. Perhaps more surprising is that six of the conflict countries managed to be classified as medium performers. These were Colombia, Iraq, Nepal, Palestine, Sri Lanka and Sudan—in most of these, conflicts have been confined to one isolated part of the country. Among the post-conflict countries, one country gained the medium classification (Bosnia and Herzegovina) and one the high (Serbia and Montenegro). These are both countries in which conflict ended some time ago. A low proportion of the conflict countries (about 30%), but a high proportion of the post-conflict countries (nearly three-quarters), were imbalanced. The nature of the imbalance was rather mixed, showing no particular pattern.

Another category of country that might be expected to make peculiar choices is that of oil countries. This category of countries has been shown to be associated with unequal income distribution, and mostly poor growth and poor basic HD in relation to a country's resources, possibly owing to various manifestations of the "Dutch Disease" (see Ranis and Mahmood 1992; also Auty 2001). This is indeed confirmed by the quite large numbers of oil countries in the low category (four out of 13), with only two (Oman and Libya) in the high category. Countries in the low category include Algeria, Angola, Nigeria and Yemen. Although these countries show quite high levels of imbalance, no systematic pattern emerges.

Turning to the transition countries, there is generally good performance, with six showing high, 13 medium, and none low performances. Only two of the 19 show political deficiency.

It is interesting to explore the extent to which high performance according to our indicators is correlated with high levels of overall "satisfaction with life" on the basis of a 0–10 ladder scale (see Table 7.6).[6] In fact, we find no evidence of a systematic relationship. In both the top and middle third of countries in terms of satisfaction with life, one quarter are classified as high, just over two thirds as medium, and just 6% fall into the low category. In the bottom third of countries, 37.5% are in the high category, half in the medium category and 12.5% in the low category. The countries with the lowest life satisfaction have more representatives in the high category and more in the low category than the top two-thirds of countries. Nor does there appear to be anything systematic about the particular dimension on which countries perform well or poorly in relation to overall life satisfaction. This contrasts with our earlier work, which showed a quite high and significant correlation across countries between life satisfaction and HDI ranking. Bringing in the broader dimensions of HD does not increase this correlation, as one might expect, but rather seems to reduce it.

[6] The data come from the World Database of Happiness and are primarily based on World Values Surveys from 1995 to 2005 in which respondents were asked to rate their satisfaction with "your life as a whole".

Table 7.6. Performance according to satisfaction with life

	Top-third countries	Middle-third countries	Bottom-third countries
Overall classification			
High	4	4	6
Medium	11	11	8
Low	1	1	2
Proportion imbalanced (%)	37.5	31.2	50.0
HD superior	2	0	2
HD deficient	0	0	0
Social superior	1	0	0
Social deficient	2	1	2
Economic superior	0	1	1
Economic deficient	1	2	0
Politics superior	0	1	0
Politics deficient	0	0	1

V. What Have We Learned About Choices?

The data do confirm what we had concluded from previous work, i.e. that not all good things always go together. Given that human development is made up of many types of freedom, capability or choice, some aspects may be promoted in some conditions and others at other times. In our classification, only eight out of 130 countries with data for all four categories were consistently categorized in the same way across categories: two as consistently high, five as consistently medium, and one as consistently low. About half the entire sample of countries showed deficiency or superiority in one category. However, the consistently weak performance of countries suffering violent conflict indicates that high priority has to be given to policies that help avoid it.

At the outset we hypothesized that alternative patterns of behavior might be dictated by political choices, by constraints, or by culture and history. Can we say more about this in light of the evidence? Our findings suggest that many poor countries are doing badly on economics and on basic HD. Despite this, a number do better on political and social aspects. Is this a matter of choice? There are three possibilities: (1) they choose to promote the social and political dimension at the expense of economic and basic HD; (2) they choose to promote the social and political despite weak economics and poor basic HD; or (3) all these developments just happened under the pressure of various external and internal forces.

It seems to us that (1) is unlikely, given the expressed desire to promote economic growth and basic HD (e.g. meet the Millennium Development Goals) and the fact that there is no obvious major resource cost in improving performance on social and political aspects. It seems more likely that weak performance on economic aspects and basic HD is a consequence of deep constraints—including weak government capacity, heavy indebtedness, and frequent violent conflicts—and not, at least with the worst performers, a matter of choice. But, given the low resource costs of the social and political aspects, some of these can be chosen even in the context of low-income economies. One needs to unravel the two categories of the social and the political to consider which aspects can be chosen and which occur exogenously.

The social dimension, as interpreted and measured here, is partly a matter of income distribution, and partly a matter of having close social and family relations and tolerant neighbors. We use the male suicide rate as an indicator to reflect how stressful life is. The income distribution variable can be influenced by the government (if with difficulty). The other variables probably could be influenced too—for example, if physical security is very low because of poor policing, social relations may be worse; and policies towards education, the media and discrimination may contribute to improving aspects of social relations. But to a considerable extent these variables are the outcome of social and economic forces, not government policy. It seems likely that these variables (particularly the ones involving relationships) depend in part on the size of communities (being stronger in rural than in urban communities), and on the spare time people have available (being stronger when people are less busy).

Our results suggest that poor countries which are socially superior are so mainly because of their superior income distribution (with data largely missing for the other indicators in the social category); for countries that are socially deficient, in contrast, the judgement is made on the basis of a wider range of social indicators.

Putting all this together suggests, very speculatively, that one might expect the social side to do better in relation to economic aspects at lower levels of urbanization and employment—i.e. at lower levels of development. This is broadly what we find, and we would argue that this is more a matter of the stage of a country's development and less of governments' or people's choices.

The political category is again a composite; it includes collective political violence, which is sometimes chosen but can result from exogenous forces. It also covers the rule of law, over which governments have some influence but which evolves slowly, with inputs from civil society as well as government; and it includes political and civil liberties, which is the one variable that can be said to be chosen, albeit, especially in the case of low-income countries, under heavy influence from the donor community. Thus, as far as this aspect of politics is concerned, the fact that some poor countries do better in politics than in other categories may be due to choices they make, not at the expense of doing well on other elements, but as something they can choose without sacrificing other aspects.

The experience of middle-income countries partly supports what has been said above, and partly indicates the wider range of choices open to middle-income countries. In the first place, many are socially deficient. This does suggest that this aspect tends to lag as development proceeds, perhaps for the reason given above—people become more urbanized and disconnected and have less time, while government efforts that might compensate for this, through policing and redistributive policies, are not always in place or effective. Sometimes, of course, such compensatory action does occur, as shown by the countries that do score high on the social category—14 out of the 77 middle-income countries. This ratio is similar to that of the low-income countries, where nine out of 50 got a high classification on the social category (though again, this was largely owing to a superior income distribution).

The middle-income countries show considerable variation in the political category, with 36 in the high category but 20 in the low category, and almost equal numbers being deficient and superior, according to our methodology. This suggests that countries make different choices in the political category, of course, heavily constrained by history. Yet the limited, or short-lived, influence of history is most clearly demonstrated by the special position of the transition countries, which are categorized in this way precisely because of their past, yet currently show a rather balanced performance. One might have expected them to be superior in basic HD, given their history, which put great emphasis on health and education, politically and possibly economically deficient, and socially mixed—good on income distribution and possibly poor on social relations. Yet this is apparently not the case.

The lack of systematic connections between life satisfaction and performance on our four dimensions of HD could be interpreted in two very different ways. One would be to argue, along with Layard (2005), that life satisfaction (or "happiness") should be the overriding single indicator of success and hence the sole objective of development. The lack of correlation with other measures of performance might be taken as a good reason for adopting this position. Alternatively, one might argue, along with Amartya Sen, that development is about expanding choices, which is better captured by our four dimensions than by a single somewhat arbitrary measure of life satisfaction; moreover, to the extent to which life satisfaction indicates that people's expectations adapt to their circumstances, it becomes a poor indicator of country performance and a faulty guide to development.[7] We are inclined to the latter view, but perceptions are also important, and consistently low views of life satisfaction are a matter which should concern decision-makers, along with our more objective indicators.

In conclusion, the many patterns of behavior indicate that while countries may be constrained by history, culture and initial conditions, they also have choices. Even low-income countries can achieve well in all categories, and even high-income

[7] Sen takes this view (1979, 1985, 1987, 1993, 2002).

countries can achieve poorly. The first gives reason for optimism, the second for pessimism. Finally, it is necessary to reiterate that our methodology is patently arbitrary and subject to refinement and checks for robustness. Our findings should therefore be viewed as suggestive only.

References

Alkire, S. (2002), *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction* (Oxford: Oxford University Press).

Auty, R. M. (2001), "The Political Economy of Resource-Driven Growth", *European Economic Review*, 45: 839–46.

Boozer, M., Ranis, G., Stewart, F., and Suri, T. (2003), *Paths to Success: The Relationship between Human Development and Economic Growth*, Economic Growth Center Discussion Paper (New Haven, Conn.: Yale University).

Camfield, L. (2005), "Researching Quality of Life in Developing Countries", *ESRC Research Group on Well-Being in Developing Countries Newsletter*, 3(1); available at <www.bath.ac.uk/econ-dev/wellbeing/news/newsletter-april-05.htm>.

Doyal, L., and Gough, I. (1991), *A Theory of Human Need* (Basingstoke: Macmillan).

Finnis, J., Boyle, J. M., and Grisez, G. (1987), *Nuclear Deterrence, Morality and Realism* (Oxford: Clarendon Press).

Freedom House (2006), "Freedom in the World 2006", available at <www.freedomhouse.org/uploads/pdf/Charts2006.pdf>.

ILO (International Labour Organization) (various years), "ILO LaborSta", available at <http.laborsta.ilo.org>.

Layard, R. (2005), *Happiness: Lessons from a New Science* (London: Penguin Press).

Marshall, M. G. (2002), *Global Terrorism: An Overview and Systematic Analysis*, Center for Systemic Peace Occasional Paper 3; available at <www.systemicpeace.org/CSPpaper3.pdf>.

Narayan-Parker, D., Chambers, R., Shah, M. K., and Petesch, P. (2000), *Voices of the Poor: Crying Out for Change* (New York: Oxford University Press).

Nussbaum, M. C. (2000), *Women and Human Development: The Capabilities Approach* (Cambridge: Cambridge University Press).

Ranis, G., and Mahmood, S. (1992), *The Political Economy of Development Policy Change* (Cambridge, Mass.: Basil Blackwell).

—— Stewart, F., and Ramirez, A. (2000), "Economic Growth and Human Development", *World Development*, 28: 197–219.

—— —— and Samman, E. (2006), "Human Development: Beyond the Human Development Index", *Journal of Human Development*, 7: 323–58.

Rawls, J. (1972), *A Theory of Justice* (Oxford: Clarendon Press).

Sen, A. K. (1979), "Utilitarianism and Welfarism", *Journal of Philosophy*, 76: 463–89.

—— (1985), "Well-being, Agency and Freedom", *Journal of Philosophy*, 82: 169–221.

—— (1987), *The Standard of Living* (Cambridge: Cambridge University Press).

—— (1993), "Positional Objectivity", *Philosophy and Public Affairs*, 22: 126–45.

—— (2000), "A Decade of Human Development", *Journal of Human Development*, 1: 17–23.

Sen, A. K. (2002), “Health: Perception versus Observation”, *British Medical Journal*, 324 (7342): 860–1.

Stewart F., and Fitzgerald, V. (eds), (2001), *War and Underdevelopment: The Economic and Social Consequences of Conflict* (Oxford: Oxford University Press).

UNDP (United Nations Development Programme) (various years), *Human Development Report* (New York: UNDP).

UN-OHRLLS (various years), “List of Landlocked Developing Countries”, <www.un.org/special-rep/ohrlls/lldc/list.htm>.

WHO (World Health Organisation) (various years), “Suicide Rates per 100,000 by Country, Year and Sex”, <www.whoint/mental_health/prevention/suicide_rates/en/index.html>.

World Bank (1999, 2000), “Levine–Loayza–Beck Dataset on Finance and the Sources of Growth and Financial Intermediation and Growth”, <http://go.worldbank.org/40TPPEYOC0>.

—— (2004), *World Bank Development Indicators 2004*, available at <http://go.worldbank.org/U0FSM7AQ40>.

—— (2006), *World Bank Development Indicators 2006*, available at <http://go.worldbank.org/U0FSM7AQ40>.

—— (2007), *World Bank Governance Indicators*, <http://info.worldbank.org/governance/wgi2007>.

World Database of Happiness (various years), <http://worlddatabaseofhappiness.eur.nl>.

World Values Survey (various years), available at <www.worldvaluessurvey.org>.

Yale Center for Environmental Law and Policy and CIESIN (2005), *Environmental Sustainability Index: Benchmarking National Environmental Stewardship* (New Haven, Conn.: Yale University).

CHAPTER 8

POVERTY AND FAMINES

AN EXTENSION

ASHUTOSH VARSHNEY

Scholars are normally elated when they produce one big idea in a given field. Amartya Sen has produced two big insights, not one, in the field of hunger and famines. One of the insights is about how democracy prevents famines; the other concentrates on the role of entitlement failures in causing famines.

The two insights work at different levels of causation. Entitlement failures—i.e. the inability of vulnerable families or groups to command food, or work, to prevent starvation—are the *proximate* cause of famines. Since such failures are more likely to arise in authoritarian political settings, the nature of the polity enters the analysis of famines at the *underlying* level of causation. "A free press and an active political opposition", says Sen, "constitute the best early warning system" (Sen 1999: 181). When these institutions, routinely present in a democratic polity, exert pressure, the government is forced to respond to early warnings. As a result, a democratic government finds a way to import, or procure, food from elsewhere, and supplies it to groups threatened by famine. Alternatively, democratic governments create emergency employment programs to generate incomes for the vulnerable groups, thereby enabling them to buy food available in the market. Entitlement failures are thus prevented from occurring, and an imminent famine does not become an actual famine.

While these two arguments are fundamentally compatible, what I wish to emphasize is that the methods used to generate them are strikingly different. The argument

about the relationship between democracy and famines is based on *variation in outcome* as a method of comparison. But the argument about the relationship between entitlements and famines is based on *commonality in outcome*, not variation. Let me explain.

The claim that famines do not take place in democracies is rooted in a dramatic *contrast* between India and China (Sen 1982). If we go by the standard yardsticks of infant mortality, life expectancy and malnourishment of children, it is clear that post-independence India has not looked after the health of its citizens well, but it has posted an enormous success to its credit: no famine has taken place in India since independence. India's last famine occurred in 1943, when the unelected British rulers were still in control of their Indian colony. Sen shows how after 1947 the institutions of a democratic polity, especially the press and opposition parties, played a role in spurring pre-emptive government action whenever the food supply dropped appreciably—for example, in the mid-1960s, in the first half of the 1970s and in 1985–7. In contrast, as reflected in its considerably better nutritional, infant mortality and life expectancy statistics, China has taken better routine care of its citizen's health, but it also witnessed one of the worst famines of the twentieth century, in which close to 30 million people died. The famine took place during the Great Leap Forward (1958–61). In the absence of a free press, good information on the scale of the tragedy was in all probability not available to the government, and since the system had neither free journalistic reporting nor opposition parties, Mao's regime felt no great pressure to take preventive action, even as millions starved to death.

In contrast, the central argument of *Poverty and Famines* (Sen 1981), a classic in the field of development,[1] was built upon four cases of famines: the Great Bengal Famine of 1943, and the famines of Ethiopia, Bangladesh and the Sahel, each of which took place in the 1970s. No *non-famines* were used for comparison. Sen argued that the conventional food availability decline (FAD) thesis could not explain the outcome. The FAD thesis had customarily linked famines to declines in the availability of food. Sen showed that the declines in food availability were either trivially small before or during these famines, or no declines had actually occurred. Entitlement failures caused those famines, as some vulnerable groups had neither the incomes to purchase food at a nutritionally critical time, nor any legally enforceable right to demand food, or employment, from public authorities.

Logically, if entitlement failure is the reason for famines, we ought to be able to demonstrate that *the opposite is true when famines do not take place*. In other words, if Sen has correctly identified the causal mechanism, we should be able to prove that when entitlements do not fail, famines cannot occur. Sen may still be right about the cause of famines, but without a comparison with non-famine outcomes, he

[1] As of 2007, November 30, according to the Google Scholar database, it is also Sen's third most frequently cited book. Only *Development as Freedom* and *Inequality Re-Examined* have been cited more often.

cannot be *sure* that the causal mechanism he has identified is indeed the right one. Studies based on commonality of outcome (or unvarying values of the dependent variable) can certainly knock down an existing theory, but as King, Keohane and Verba (1994) have argued, such studies cannot establish a new causal explanation.

It follows that in order to be sure that the causal mechanism has been correctly identified, we need to do with *Poverty and Famines* what Sen did with the India–China comparison. We need variation in research design. In what follows I do this briefly by bringing in some well-known cases of non-famines. The argument that emerges supports Sen at the proximate level, but at the underlying level of causation it adds a new variable: technological dynamism in agriculture. An implication is that democracy is not the only variable that protects entitlements. Technological dynamism in agriculture can also lead to the same result by more or less obviating the need for entitlement protection.

Before I develop my argument in detail, let me begin with a methodological appraisal of *Poverty and Famines*. It may not have had variation built into its design, but Sen's research strategy still had some parts that were exemplary (section I). Having highlighted those strengths, I will bring non-famines into Sen's design (section II), and draw appropriate conclusions. Finally, in conclusion (section III), I will note how Sen, in some of his works published after *Poverty and Famines*, briefly—and intriguingly—anticipates my argument, though he does not develop it fully.

I. Causal Mechanisms

Published in 1981, *Poverty and Famines* had the customary traits associated with Sen's previous scholarship: mastery of logical reasoning, surgical economy of expression, and intense engagement with a problem of profound human importance. But *Poverty and Famines* also had something strikingly novel. Until then, Sen had never conducted in-depth case studies, nor used case studies as a tool of theory formation. Although Sen had written some empirical papers before, especially on agricultural economics (Sen 1966, 1976), his arguments were better known either for abstract reasoning, or for addressing the conceptual and normative issues that empirical matters often raise: for example, how to conceptualize inequality before measuring it, or how to measure poverty (Sen 1973).

Case studies are known for what has come to be called in non-economic social sciences "process tracing" (George and Bennett 2005). Empirical in spirit, process tracing is a step-by-step account of *how* the outcome in which we are interested—starvation and famines in this case—developed. The cases may be

selected for theoretical reasons, but the investigation is deeply empirical and process-oriented.

That is partly why *Poverty and Famines* carried such weight beyond economics. For the Bengal famine of 1943, Sen wrestled with the archives, and for the more contemporary famines, he immersed himself in the existing case materials. An exploration of the causes of famines required a systematic investigation of *how* exactly a famine developed without the superimposition of theory. Though Sen used his well-known capacities for abstract and philosophical reasoning to elaborate the concept of poverty and entitlements, his explanation of the causes of famines was ultimately inductive. Arguments about famines had not previously witnessed such a blending of philosophical acuity and inductive thoroughness.

Case studies, Sen's preferred method in *Poverty and Famines*, are often contrasted with "large-N" research, which studies a lot of cases together, or draws inferences from a large number of observations. The latter is notorious for mixing up causes and consequences. At its best, large-N research can establish *causal effects*, but it cannot give us an account of *causal mechanisms* (Gerring 2007). In contrast, by clearly laying out what came first and what happened later, process tracing is better at establishing causal mechanisms, though it may not be a method for establishing causal effects. For causal mechanisms to be investigated in depth, the number of cases must be reduced to a small-N research design. Discovering causal mechanisms requires empirical intimacy with case materials. Large-N research does not allow intimacy, only fleeting acquaintance.

In short, by leveraging one of the greatest strengths of case studies—identification of causal mechanisms—*Poverty and Famines* highlighted a serious flaw in the conventional wisdom about famines. According to the FAD thesis, famines are caused by a decline in aggregate food supply, or availability. If food availability remains unchanged or goes up, starvation, or famines, will not take place.

This formulation manifestly provokes a question about mechanisms. Overall food supply is an aggregate measure; starvation affects particular groups, not all. Therefore, we need an account of how something happening at an aggregate level can affect a fraction of those covered by the aggregate description. The opening lines of *Poverty and Famines* identify this problem with arresting immediacy: "Starvation is the characteristic of some people not having enough food to eat. It is not the characteristic of there being not enough food to eat" (Sen 1981: 1).

I.1 Entitlements

Why do some people not have enough food to eat? It is in answering this question that Sen developed the concept of "entitlements".[2] The concept of entitlement is

[2] While entitlement as a concept was first used in *Poverty and Famines*, it is in a later book that, in my view, it acquired radical clarity (Drèze and Sen 1989).

legal, not normative: "The notion of entitlement . . . must not be confused with normative ideas as to who might be 'morally entitled' to what. The reference instead is to what the law guarantees and supports" (Drèze and Sen 1989: 22–3).

In a market-based economy, entitlements can take various forms, three of which are critical: direct entitlements, exchange entitlements, and public policy-induced entitlements. When peasants produce food crops for family consumption, we speak of food as their direct entitlement. But one does not have to produce food crops to acquire food. Food can be acquired in a market exchange. Sen speaks of such exchanges as exchange entitlements. Finally, food can also be made available by targeted governmental supply; alternatively, governments can create public employment programs, which give those participating in such programs an income, which in turn can be used to buy food. This is an example of how public policy can protect entitlements.

Why do entitlements fail? Process tracing allows Sen to demonstrate that unless we employ a disaggregated approach and look at how different groups are affected, we cannot answer the question. Nothing very insightful can be inferred from aggregate statistics such as food availability in a country or a province.

Since most famine deaths occur in the countryside, it is analytically useful to think of the various groups that make up the countryside. The rural population can be divided into the following groups: those who own land—a lot of it (large farmers or landlords), or just a little (small peasants); those who own only their labor but no land (large landless laborers); those who do not own land but do not have to hire their labor out, seeking instead a living with a share of the crop (sharecropping tenants); those who have skills with which they make a living without working land (barbers, carpenters), etc.

Once the group characteristics are laid out, one can begin to see how entitlement failures may emerge. If the crop of those producing enough food for themselves is wiped out for some reason—because of a drought or flood, for example—their direct entitlements fail. If the government does not provide employment or food, exchanging what they own for food—land, cattle or labor—will be their best alternative. Given how often exchange is what the poor are left with, the concept of exchange entitlement becomes the centerpiece of Sen's analytic scheme. Simply put, exchange entitlement signifies the set of commodities that a person can acquire in return for what that person owns. Labor can be exchanged for a wage; wages can be exchanged for food. A carpenter can put his skill into a product, which he exchanges for a price, and that can in turn be exchanged for food.

I.2 Famines: Reconstructing Two Examples

The FAD view looks at overall production and assumes consequences without factoring in the different natures of group endowments and entitlements. By

examining the causes of famines, Sen poses the issue very starkly. Can famines take place even when nothing happens to food production? Can thousands, even millions, of people die of starvation even when food production declines only marginally, not significantly? If the answer is yes, the argument will have to be formulated in terms of distribution of endowments and entitlements, not in aggregate terms.

To find out whether that is true, Sen, as noted above, studied four famines: the Great Bengal famine of 1943, in which an estimated 3 million people died (Sen 1981: 52); the Ethiopian famine of 1973/4, in which anywhere between 50,000 to 200,000 people lost their lives (Sen 1981: 86); the famine in the Sahel (1973), whose death toll was at least 100,000 (Sen 1981: 116); and the Bangladesh famine of 1974, for which the mortality estimates range from 26,000 to 100,000 (Sen 1981: 134). The Bengal and Ethiopian famines illustrate the causal mechanisms best.

Of the four famines, the Bengal famine of 1943 has attracted the most attention (Arrow 1982; Bose 1990; Bowbrick 1986; Goswami 1990; Islam 2007; Ravallion 1997; Sen 1987). Part of the reason is the sheer scale of mortality. But the more important reason was that Sen identified it as a "boom famine" (Sen 1981: 75). The food supply in Bengal in 1943 might have been lower than in 1942, but 1942 saw an unusually high output. If anything, food output in 1943 was 13% higher than in 1941, which was not, however, a famine year. If the food supply in 1941 was so much lower and still a famine did not take place, then the causes of the 1943 famine would lie elsewhere, not in food output. Sen showed that sharp movements in the exchange entitlements of some groups constituted the primary cause. Peasant cultivators and sharecroppers were affected least. Landless agricultural laborers, fishermen, transport workers and paddy huskers were the biggest victims. They did not directly own land and its products, but bought food in the market. Once a war economy had created enormous inflationary pressures in the economy, food became so expensive that these groups simply could not buy enough food for their families. Mass starvation was the result.

The Ethiopian famine, unlike the Great Bengal famine, is not normally commented upon, or not enough in any case, but it illustrates the causal mechanisms in a very clear way, and is worthy of brief reconstruction.

During 1972 to 1974, the north-eastern province of Wollo was hit hard by a shortfall in rain. The total output in Ethiopia, however, did not drop significantly: it went down by a mere 6 to 7%. But the death toll was not less than 50,000, and possibly as high as 200,000, mostly concentrated in Wollo.[3] In other words, while Wollo was doing badly, much of Ethiopia was going through a normal food cycle.

Sen's process tracing allows him to identify what can be called the three interconnected puzzles of the Wollo famine. The resolution of these puzzles shows how the famine developed.

[3] The provinces of Tigrai and Hareghe were also affected.

1. Food from other parts of Ethiopia did not reach Wollo. Why not?
2. If food from other parts did not reach Wollo, one would expect food prices to rise, for demand for food would exist but supplies, in the absence of replenishments, would decline. Food prices, however, did not go up. Why not?
3. If food prices did not go up, one would expect food to be affordable. People, however, starved to death. Why?

Yet again, Sen convincingly shows that if we focus on the aggregate food supply, we will not be able to resolve these puzzles. But if we start looking at who the famine victims were, we begin to get answers. Small farmers were the greatest victims.

As the output drastically declined in Wollo, small farmers lost their normal and primary source of income. Since their incomes failed, the effective demand for food also declined, even though a need for it existed. Food prices, therefore, did not rise. As output dropped, many farmers tried to sell their lands. With very few takers, land prices plummeted. Some farmers also tried to sell their livestock. Some livestock, of course, had died, but others could not be sold. With falling incomes, livestock prices also fell dramatically. Further, as the farmers' livelihood declined, so did that of the artisans, who depended for their welfare on the welfare of the land owners. Finally, since the effective demand for food fell so drastically, there was no market incentive for grain traders outside Wollo to send in supplies. The normal operations of market processes ensured mass starvation.

The four case studies allowed Sen to derive two big conclusions. First, famines are entitlement failures. At best, food supply is one of the factors, but it is by no means the most important one. Second, not all groups suffer from famine. Certain groups are especially vulnerable to famine threats. What one owns and what it can be exchanged for is critical. On the whole, those owning labor are more vulnerable than those owning land; and those owning small patches of land are more vulnerable than larger farmers. Unless the government can provide supplementary food or employment in times when a famine is approaching, market processes, left to themselves, can actually produce mass starvation.

II. The Green Revolution and Famines

While the above argument flows neatly from the cases chosen, the Green Revolution, which transformed agriculture, especially in Asia, is missing in *Poverty and Famines*. Once brought into analytic gaze, it offers revealing contrasts, comparable to the radical sharpness of Sen's India–China contrast on famines. Ironically, the Green Revolution reinforces Sen's argument about entitlements, but it produces a new candidate—technological dynamism in agriculture—for what prevents

entitlement failures. In addition to democracy, technological dynamism in agriculture emerges as an underlying factor pre-empting entitlement failures.

It is noteworthy that all of the 1970s famines studied by Sen occurred in regions that did not go through a green revolution: Sub-Saharan Africa and Bangladesh. For reasons outlined below, while the agrarian ecologies of Sub-Saharan Africa have made virtually the entire continent resistant to the Green Revolution, something that continues to be true even today, most of Asian agriculture had been technologically transformed by the early to mid-1970s. Bangladesh, where a famine did take place in the 1970s, was among the few exceptions in Asia.[4]

The key question is: why did famines *not* occur—in the 1970s or later—in those parts of the world which went through a green revolution? To be more specific, why did famines occur in Sub-Saharan Africa and Bangladesh, but not in India's Punjab, Malaysia and Indonesia?[5] As is widely known, the mid-1970s were generally a difficult time for world agriculture, partly because of the weather in several parts of the globe and partly because of the oil shocks. Due to the energy intensity of the Green Revolution, the oil price increases led to huge cost escalations after 1973, but the Green Revolution areas nonetheless emerged relatively unscathed and without nutritional catastrophes. In contrast, Bangladesh and Sub-Saharan Africa suffered famines.

Let us first briefly go through the technological aspects of the Green Revolution.[6] The new technology, as it was called then, had two analytically separable sides: the biochemical (seed, water, fertilizer) and the mechanical (tractors, threshers and combine harvesters). The core of the revolution was biochemical. Once agricultural scientists had developed the new high-yielding variety (HYV) seeds in the early 1960s, piloted their use on experimental farms and confirmed by the mid-1960s that the new seeds raised yields significantly, it also became clear that for best results, the new seeds required (1) timely and controlled quantities of water, and (2) an optimal dosage of chemical fertilizer. Only irrigation could make (1) possible, not unpredictable rains. The HYV seeds, irrigation and chemical fertilizer thus became the yield-increasing core of the Green Revolution.

Though there are several reasons why Africa did not have a green revolution, the absence of irrigation was the key factor, as the father of the Green Revolution has recently reminded us (Borlaug 2007). Sub-Saharan Africa was, and has remained, heavily dependent on rain; very little of its cropped area is irrigated. In contrast, a substantial part of Asian agriculture was irrigated as early as the 1970s. Bangladesh

[4] The other Asian exceptions in the 1970s were the Maoist countries China, Vietnam, Cambodia and Laos.

[5] The only pre-1970s case study in *Poverty and Famines* is functionally equivalent: the 1943 famine took place in Bengal, which was not at the technological edge of Indian agriculture at that time. After the British invested in irrigation systems, Punjab established a huge technological lead over the eastern part of the country. Punjab did *not* have a famine in 1943; Bengal did.

[6] For more details, see Varshney 1995 (10–14, 52–7).

was then a major Asian exception. It was flood-prone, not drought-prone like Sub-Saharan Africa, but its agriculture in the 1970s suffered from a lack of substantial irrigation capacity. The country normally had an excess of rain, whereas the Green Revolution, due to the requirements of the new seeds, required controlled and timely irrigation. Bangladesh's agrarian ecology thus allowed only a slow adoption of the new technology. A green revolution did finally take place, but only later, in the 1980s, when a serious change in agricultural policy came about. The World Bank notes:

> Most of the increase in rice production since 1980 has been on account of the winter season (January–May) *boro* crop. Traditionally, the main rice crop in Bangladesh was the monsoon season *aman* crop (August–December), which depends almost entirely on rainfall and regular seasonal flooding of rivers and streams. *Following the liberalization of imports of diesel engines and pumps for tubewell irrigation and expansion in fertilizer in the 1980s and the growing use of high yielding varieties of rice . . .*, the area planted to *boro* increased sharply. The *boro* area has almost quadrupled since 1980, replacing the lower yielding *aus* rice crop (April–August) in many areas. In the meanwhile, as a result of concomitant productivity improvements, *boro* production has increased almost six-fold.
>
> (World Bank 2007a: 57; italics added)

What relationship between the Green Revolution and famine prevention can be proposed? Consider first the direct and exchange entitlements of small farmers owning lands in the so-called Green Revolution belts. I shall examine later the larger implications for those parts of an economy not touched directly by the Green Revolution.

It is clear, at least in retrospect, that the mechanical inputs of the Green Revolution had a size bias: only the bigger farmers could afford tractors or combine harvesters. But the biochemical inputs were relatively size-neutral (for details, see Varshney 1995: 121–33). Only small capital outlays were required to buy seeds and chemical fertilizer in the marketplace and to purchase water through an irrigation channel. It was not surprising, therefore, that small farmers also benefited tremendously from the Green Revolution in highly irrigated agricultural zones or countries. In 1962/3, farms smaller than five acres were unable to produce a wheat surplus in Punjab, the heartland of India's green revolution, but after the introduction of the Green Revolution technology, a surplus could be produced on farms as small or marginal as 1.25 to 2.5 acres (Blyn 1983; Chaddha 1986). A massive technological breakthrough, via productivity enhancement, thus protected small-farmer entitlements. Further, by reducing the length of the cropping cycle, the new seeds also made multiple cropping possible.

In the current stage of research on agrarian political economy, the focus is rightly on the exhaustion of the Green Revolution by the late 1980s and on how to deal with the second-generation effects of the excessive, rather than optimal, use of Green Revolution technology: overuse of fertilizers and pesticides and the increasing

toxicity of soil and water in some agronomic regions of the world; salinized upper soils and declining soil fertility in some other parts; and sinking groundwater levels and the prohibitive costs of small irrigation in still other parts (World Bank 2007b: ch. 8). Today, we have a different scientific task ahead of us: namely, how to readapt, or reinvent, Green Revolution technologies for the next generation of challenges. But it is worth recalling the transformative role played by the Green Revolution four decades back. In the 1960s, 1970s and 1980s, the Green Revolution vastly improved agricultural productivity in Asia, increasing grain supplies and making food security possible for millions who would otherwise have been quite vulnerable.

As the examples from India, Malaysia and Indonesia below will illustrate, there are two principal ways in which the Green Revolution substantially reduced mass hunger in Asia and prevented famines:

(1) by making small farmers in irrigated areas more productive;
(2) by making it possible for governments to accumulate buffer stocks, which could be used:
    (a) to stabilize food prices all over the country, including prices in those parts that were not irrigated and could not use the Green Revolution technology directly, and
    (b) to make food available in times of supply shocks.

A third, macroeconomic, route to greater welfare is also conceivable, though the full development of the idea need not detain us. The Green Revolution made quick industrialization possible in several cases, especially in South-East Asia. As we have known since Marshall, industrialization requires food and labor surpluses. It is also normally accompanied by an eclipse of agriculture, but the eclipse is least painful when it is led by a productive agricultural sector, releasing labor and food surpluses (and also possibly savings) simultaneously. Industrialization based on productive agriculture proceeds much better than industrialization based on repression of the agricultural sector, as it creates gainful employment for those no longer necessary in agriculture (Varshney 1995: ch. 2).

II.1 India

From 1965 to 1967, two successive droughts brought India's food production down to the level of 1956/7. Public food stocks, for feeding urban India's poor and lower middle classes, were almost entirely dependent on subsidized grain imports from the United States, which reached a staggering 10 million tons in 1975/6 (Varshney 1995: 48–9). To explain where India was headed, a popular book of the mid-1960s argued that those wounded in the battlefield were normally split into three camps: those only slightly wounded, who were not a cause for concern; those more seriously wounded, who could be saved through surgery; and those so grievously

injured that they were left to die. India was argued to be in the third category: "No matter how one may adjust present statistics . . . it will be beyond the resources of the United States to keep famine out of India during the 1970s" (Paddock and Paddock 1967: 217).

Defying these pessimistic predictions, India's food production started rising after 1967/8, as a green revolution swept through Punjab, Haryana and western Uttar Pradesh. In 1965/6 and 1966/7, India's grain production was 72.3 and 74.3 million tons respectively. By 1970/1, India was producing 95 million tons of grain. Moreover, in 1971 India unilaterally terminated American wheat imports, as public stocks of domestically procured grains had reached 8.1 million tons. Further, by 1986/7, India's public stocks had risen to 30 million tons, which ensured that the acute drought of 1987/8 did not turn into a famine. Until 1965/6, most of the increases in food production were due to acreage expansion. But for roughly three decades after 1966–7, productivity per acre, not increases in cultivable acreage, drove the growth in food production, thus lifting India from the agricultural abyss of the mid-1960s.

The Green Revolution had two effects on the entitlements of India's poor citizens, one direct and the other indirect. The best example of the direct effect was noted above. Even though it was initially believed in some quarters that only the bigger farmers would benefit from the Green Revolution, smaller farmers also soon started producing food surpluses on their meager farmholdings. That gave them greater capacity not only to feed themselves, but also to withstand shocks. The biochemical inputs could be easily purchased, and the mechanical inputs rented, if not bought.[7]

The indirect effect had to do with how the Green Revolution enhanced the buffer stocks of food grain, allowed the distribution of greater quantities to the poor, and moderated food prices overall. Consider the droughts of Maharashtra from 1970 to 1973. Food at a reasonably low price could be moved to Maharashtra due to the rising productivity and the surpluses produced, in Punjab, Haryana and western Uttar Pradesh. To prevent famine, Maharashtra at the peak of the crisis was providing employment to as many as 5 million people through public works programs (Drèze and Sen 1989: 129).[8] Had the Green Revolution surpluses not been available, food imports on international commercial prices would have made India's external macro-balances extremely precarious. Unlike now, India's foreign exchange reserves were meager in the early 1970s. Exports were not rising and the dollar remittances from Indian migrants were yet to fill the public treasury. In short, in addition to the institutions of democracy, especially the free press and opposition parties, exercising pressure on the Maharashtra government, as Sen has noted, the

[7] While the welfare implications of the Green Revolution for small farmers in irrigated areas were significantly positive, the implications for the landless were ambiguous (Blyn 1983; Varshney 1995: 133–8).

[8] See also Echeverri-Gent (1993).

buffer stocks made possible by the Green Revolution were almost certainly another reason for the absence of famine in Maharashtra from 1970 to 1973.

II.2 Malaysia and Indonesia

In South-East Asia, too, the Green Revolution transformed agriculture, improved productivity and nutrition, including that of small farmers, and made the possibility of a famine, even in times of production downturns, increasingly remote. The evidence for these claims is contained both in anthropological village-level accounts and the larger aggregate analyses.

In the Muda region of Kedah, a relatively poor state of Malaysia, double-cropping began in 1970, as a major irrigation scheme made the adoption of Green Revolution technology viable. James Scott notes that

> [b]efore double cropping, one third of the farm households in the region rarely grew enough rice for the annual subsistence needs of the family. . . . Double cropping in this respect has been a great boon. Even smallholding tenants with a single *relong* (.71 acre) can now grow enough rice at least to feed a family. . . . *It is a rare peasant these days who does not eat rice twice a day*. The provision of irrigation water and the use of fertilizer not only raised yields but has also made those yields more reliable, season by season. . . . the prospect of going without rice has been largely removed from the fears of even poor villagers.
>
> (Scott 1985: 66; italics added)

Furthermore, the incidence of nutrition-related diseases markedly declined, and "between 1970 and 1976 the rate of infant mortality in the Muda region was cut by nearly a half" (Scott 1985: 67).[9]

If India had a democracy at the time of its green revolution and Malaysia a semi-competitive electoral system heavily inclined towards the countryside, where most Malays lived, Indonesia's green revolution had no such confounding variables explaining the decline of mass hunger and elimination of famines. Under Suharto, Indonesians knew no political freedoms, but the countryside of what was arguably the poorest country in the world in the mid-1960s, when Suharto took power in a coup, was dramatically transformed in the 1970s. In 1965, Indonesia's daily caloric supply was 1,742 kcals, which was below the daily caloric supply not only in Malaysia and Thailand, but also in India, Pakistan and what later came to be called Bangladesh. By 1985, Indonesia's daily caloric supply was second in South-East Asia only to Malaysia's, and had left all of South Asia behind (Timmer 1991: 148). The key was a set of policies that combined government support of crop prices, rural

[9] One should, however, note that with the introduction of massive mechanization in the mid-1970s, some of the gains made by the poor in the first flush of the Green Revolution were reversed, as labor was displaced (Scott 1985: 74–81). However, the displaced labor was soon absorbed into the rising industrial sector in Malaysia. For the impact of markets on famines in general, see Ravallion (1987).

public works and new Green Revolution technology, leading among other things to higher labor productivity, benefiting the poor directly (Timmer 1991: 125).

Indeed, the food security of Indonesia was transformed so radically by the mid-1990s that when the post-Suharto shock came in 1998, and the economy shrank by a whopping 18–20% in a year, there was considerable misery all around for the poor, but no famines occurred. An economic decline of such magnitude in one year could have been expected to lead to mass starvation. But more than two decades of technological dynamism in agriculture had produced greater food security throughout the country.

In short, if entitlement failures were prevented in India and Malaysia by the combined effect of technological dynamism in agriculture and political competition, in Indonesia, which lacked democracy till 1999, the pure effect of the Green Revolution technology stands out. Technologically induced productivity and income gains, achieved over three decades of the Green Revolution, were the underlying causes of Indonesia's success in preventing entitlement failures right through the 1970s, 1980s and 1990s, but most especially during the huge economic contraction of 1997/8.

It is possible to misread my argument about the relationship between technological dynamism and famines. Let me therefore provide some final clarifications. First, my argument does not posit a sharp contradiction between markets and public action. In retrospect, it is clear that the success of the Green Revolution was premised upon a supportive government policy framework. Government intervention was quite necessary for input provision, research and extension, grain procurement, and stabilization of food prices. Markets were used, but in a larger framework which included government intervention. Second, my argument should not be read as a restatement of the FAD hypothesis. Technologically induced increases in supply are fundamentally different from increases, or declines, in food availability in a technologically stagnant framework. Moreover, the Green Revolution technology directly increased the capacity and entitlement of the smaller farmers: they did not have to depend on aggregate supplies to protect their entitlements. The FAD theory is purely aggregative.

III. Concluding Remarks

Though *Poverty and Famines* does not deal with the Green Revolution and its implications for famine prevention, my principal argument in this essay is that even if we include the Green Revolution cases, Sen's theory of entitlement holds. What changes is simply a possible line of causation at the underlying level. Sen's focus has been on democracy as one of the key underlying variables preventing

famines, but one can also show that technological dynamism in agriculture can lead to prevention of famines in both democratic and undemocratic settings. In short, technological dynamism merits a separate causal status of its own.

It should be noted that in works published after *Poverty and Famines*, Sen partially anticipates the core aspects of my argument. Though his focus while dealing with hunger has never been on the Green Revolution, his distinction between the short-run and the long-run causes of famine does take him briefly towards the following insight:

> The protection of entitlements in the short run has to be contrasted with the general promotion of entitlements in the long run. In the short run, famine prevention is essentially a question of encountering an immediate threat of entitlement failure for vulnerable groups. In the long term, of course, much more is involved, and a durable enhancement of invulnerability requires promotional policies, such as the expansion of general prosperity, the reduction of insecurity through economic diversification, and the creation of secure earning arrangements. (Drèze and Sen 1989: 66)

In another brief passage written a decade later, Sen speaks in a similar vein:

> In preventing famines, it helps to have a more opulent and growing economy. Economic expansion typically reduces the need for entitlement protection. This is a lesson of obvious importance for sub-Saharan Africa, where the lack of overall economic growth has been a major underlying source of deprivation. The proneness to famines is much greater when the population is generally impoverished and when public funds are hard to secure.
>
> (Sen 1999: 175)

My principal argument in this essay began with methodology, but its substantive essence is, for all practical purposes, an elaborate extension of this insight.

References

Arrow, K. (1982), "Why People Go Hungry?", *New York Review of Books*, 29(12) (15 July).

Blyn, G. (1983), "The Green Revolution Revisited", *Economic Development and Cultural Change*, 31: 705–25.

Borlaug, N. (2007), "A Green Revolution for Africa", *Wall Street Journal*, 26 October.

Bose, S. (1990), "Starvation Amidst Plenty", *Modern Asian Studies*, 24: 699–727.

Bowbrick, P. (1986), "The Causes of Famines: A Refutation of Professor Sen's Theory", *Food Policy*, 11: 105–24

Chaddha, G. K. (1986), *The State and Rural Transformation: A Study of Punjab* (New Delhi and Newbury Park, Calif.: Sage Publications).

Drèze, J., and Sen, A. (1989), *Hunger and Public Action* (Oxford: Clarendon Press).

Echeverri-Gent, J. (1993), *The State and the Poor* (Berkeley and Los Angeles: University of California Press).

George, A., and Bennett, G. (2005), *Caste Studies and Theory Development in the Social Sciences* (Cambridge, Mass.: MIT Press).

Gerring, J. (2007), "Case Studies", in C. Boix and S. Stokes (eds), *Oxford Handbook of Comparative Politics* (Oxford: Oxford University Press), 90–122.

Goswami, O. (1990), "The Bengal Famine of 1943", *Indian Economic and Social History Review*, 27: 445–63.

Islam, M. M. (2007), "The Great Bengal Famine and the Question of FAD Yet Again", *Modern Asian Studies*, 41: 421–40.

King, G., Keohane, R., and Verba, S. (1994), *Designing Social Inquiry* (Princeton, NY: Princeton University Press).

Paddock, W., and Paddock, P. (1967), *Famine 1975! America's Decision Who Will Survive* (Boston: Little, Brown and Co).

Ravallion, M. (1997), "Famines and Economics", *Journal of Economic Literature*, 35: 1205–42.

—— (1987), *Markets and Famines* (Oxford: Clarendon Press).

Scott, J. P. (1985), *Weapons of the Weak* (New Haven, Conn.: Yale University Press).

Sen, A. K. (1999), *Development as Freedom* (New York: Alfred Knopf).

—— (1987), "Reply: Famines and Mr. Bowbrick", *Food Policy*, 11: 125–32.

—— (1982), "How is India Doing?", *New York Review of Books*, 29(20) (16 December).

—— (1981), *Poverty and Famines* (Oxford: Clarendon Press).

—— (1976), "Famines as Failures of Exchange Entitlement", *Economic and Political Weekly*, special no. (August): 1273–80.

—— (1973), *On Economic Inequality* (Oxford: Clarendon Press).

—— (1966), "Peasants and Dualism with or without Surplus", *Journal of Political Economy*, 74: 425–50.

Timmer, C. P. (1991), "Agricultural Development and Poverty Alleviation in Asia", in C. Peter Timmer (ed.), *Agriculture and the State* (Ithaca, NY: Cornell University Press), 123–55.

Varshney, A. (1995), *Democracy, Development and the Countryside* (Cambridge and New York: Cambridge University Press).

World Bank (2007a), *The Bangladesh Paradox: Does Governance Matter to Growth?* (Washington, DC: World Bank).

—— (2007b), *World Development Report 2008: Agriculture and Development* (New York: Oxford University Press).

This page intentionally left blank

PART II

GENDER AND THE HOUSEHOLD

This page intentionally left blank

CHAPTER 9

ENGAGING WITH SEN ON GENDER RELATIONS

COOPERATIVE CONFLICTS, FALSE PERCEPTIONS AND RELATIVE CAPABILITIES[†]

BINA AGARWAL

> I do see myself, in part, as a feminist economist, in addition to having other descriptions to which I respond. This is partly because of my direct involvement with gender-related issues, but also because of my conviction that the perspective of gender inequality gives us real insight into asymmetries and deprivations of other kinds as well.
>
> Amartya Sen in conversation with the author and two colleagues (Agarwal *et al*. 2006: 351)

Amartya Sen's work has inspired and challenged scholars across many disciplines, and it has been crucial to the development of several aspects of feminist economics

† I am grateful to Kaushik Basu for his helpful comments on an earlier version of this paper.

and gender analysis. Many of his writings address gender concerns directly, but even when they do not, his work focuses on themes that are central to feminist analysis not only in economics but across the social sciences, as well as philosophy. His humanitarian approach to the subject has also encouraged many to ground their economic analysis in real life, to question narrow notions of economic rationality, and to recognize the importance of qualitative factors in determining economic outcomes, nothwithstanding their uneasy fit with formal economic models.

Over the years I have engaged with and learnt from many aspects of Sen's writings, in some cases applying or extending his ideas, in other cases contesting them. In this chapter, I bring together three forms of such engagement, each of which relates to an aspect of intra-household gender relations. In the first I apply his work on cooperative conflict within families, and his entitlement approach to famines, to explain why families might break up during famines and what this reveals about intra-family gender relations. In the second I contest his claim that women in traditional societies suffer from false perceptions about their self-interest and so become complicit in perpetuating gender inequalities. And in the third I use the context of domestic violence to demonstrate the importance of relative capabilities (and not just absolute capabilities) in determining well-being outcomes. In the process, the chapter also highlights the critical role played by a much neglected factor, women's property status, in determining their economic and social well-being.

I. Intra-Household Bargaining, Entitlements and Famines

Few contexts illuminate intra-household gender relations as starkly as extreme calamities such as famines. Why do families break up during famines? In particular, why do we commonly observe the abandonment of women and children? Although Sen does not specifically address this issue in his seminal writings on famine, at least two aspects of his work provide the basis for an answer: one, the conceptualizing of intra-household relations as relations of cooperative conflict in which outcomes depend on each person's bargaining power; the other, his entitlement approach to famine and starvation. In Agarwal (1990) I applied these concepts to the Great Bengal Famine of 1943 and came up with rather different explanations for family dissolution than those provided by two leading anthropologists—Paul Greenough and Arjun Appadurai.

I.1 Cooperative Conflict and Bargaining Power

Adapting John Nash's (1950, 1951) game-theoretic formulation of cooperative "bargaining problems" to intra-household relations, Sen (1990) conceptualizes the family as a bargaining unit in which interactions between family members are characterized by both cooperation and conflict. Household members cooperate insofar as cooperative arrangements make each of them better off than non-cooperation. Many cooperative outcomes are possible that are beneficial to both parties compared with non-cooperation, but each party has strictly conflicting interests in her/his choice among the set of efficient cooperative arrangements. Which cooperative solution emerges—for instance, how household goods and services get allocated—depends on the relative bargaining power of each member. A member's bargaining power, in turn, depends on the strength of her/his fallback position should cooperation fail (though undeniably love and concern could impinge on this in unexpected ways). This conceptualization is in contrast to the unitary model of the household *à la* Gary Becker (1981), which assumes that all household resources and incomes are pooled, and intra-household allocations depend on decisions made by the household head, who represents the members' tastes and preferences and seeks to maximize household utility.[1]

On the basis of Sen's formulation it can be argued that a calamity which leads to a collapse of the woman's fallback position and her ability to contribute to the family's material well-being (as can happen during a famine), while the husband's fallback position is sustained to some extent, could weaken her bargaining power even to the point where the man finds non-cooperation more beneficial than cooperation, creating a tendency toward family disintegration and the abandonment of women and children.

Key to understanding such outcomes are the factors which might affect a person's fallback position and bargaining power within the family in relation to subsistence needs such as food. In his entitlement approach to famine, Sen (1981) highlights two factors as significant in determining a person's ability to command subsistence: endowments (what a person owns, such as assets and labor power) and exchange entitlement mapping—the exchange possibilities that exist through production and trade, and which determine the consumption set available to a person with given endowments. Typically for rural families, the most important endowments are arable land and the ability to labor on which Sen has focused. However, I would like to suggest five additional factors. Three are in the nature of entitlements that do not derive from private ownership or market exchange, namely traditional rights in communal resources, traditional social support systems, and support from the state and civil society. Impinging on all these factors are two others—social norms embodying accepted notions about the division of labor and resources, and social

[1] Sen's conceptualization is in line with a range of efforts to provide an alternative to the unitary conceptualization of the household (for details, see Agarwal 1997).

perceptions about contributions, needs and abilities (and therefore about who deserves what). We might suggest that these seven factors would individually and interactively affect a person's ability to fulfill subsistence needs outside the family. The premise here is that the greater a person's ability to survive physically outside the family, the greater will be her/his bargaining power over subsistence within the family. Inequalities among family members in respect of these factors will place some members in a weaker bargaining position relative to others. Gender is one significant basis of inequality.

It needs to be emphasized that the bargaining framework does not require an explicit process of bargaining. Social norms that limit women's work mobility, education, or property access, or which otherwise undermine her capabilities and functionings,[2] can give men considerable power even without explicit discussion. It is also notable that many of the factors which determine a person's fallback position also influence her ability to contribute toward family well-being. If a woman loses her job and her assets, it both worsens her fallback position and diminishes the income she can bring into the family. This dual effect can undermine the benefits of cooperation that the husband might derive within marriage and so lead to marriage dissolution in periods of severe crisis, such as famine. Here, by going beyond the Nash bargaining framework, we can understand why cooperation might fail, and that the possibility of breakdown is real and not just a threat that influences the outcomes of bargaining.

Below I briefly outline the evidence of family dissolution and how this has been explained by some scholars, before presenting an alternative explanation.

I.2 Family Dissolution

Famines starkly pose the economic and moral dilemmas of intra-household food sharing and hold up a mirror to family relations that few other contexts can. The Bengal famine of 1943 provides a graphic case study in this respect. A range of evidence points to the specific disadvantages suffered by women and children of poor households during the famine.

First, among those made destitute in rural Bengal, young and middle-aged females were dominant: in January 1943, 55% of all destitutes, and 66% of those in the 15–50 age group, were women (Mahalanobis *et al.* 1946). In this age group, women destitutes also outnumbered male destitutes two to one, between January 1943 and May 1944. Children (girls and boys) in the 5–15 age category were next in number. Occupationally, the largest number of destitutes were fisherfolk, transport workers, agricultural laborers, paddy huskers and craftsmen (Chattopadhyay and Mukerjee

[2] Sen conceptualizes capabilities as the abilities or opportunities a person has to do or to be what she values, and functionings as that which she actually manages to do or to be.

1946: 18). Paddy husking was among the few occupations in which poor women, in normal circumstances, found employment.

Second, those who came to state relief centers during the worst months of the famine and were able to pay cash for food were mostly male adults, while those dependent entirely on gratuitous relief tended to be females (mostly adults), 84% of whom needed gratis aid compared with 43% of the male destitutes (calculated from Greenough 1982: 190).

Third, although excess mortality (the absolute and proportionate increases in mortality attributable to famine) was higher among males than females, the estimated absolute number of famine deaths during the critical year of 1943, and the absolute increase in mortality rates attributable to the 1943 famine, were greater for women than men in the 20–40 age group (Greenough 1982: 311). This is notable, given that childbirth and childbearing risk tends to decline for women of this reproductive age group due to famine-induced fall in birth rates. Also, most of these women were within marital relationships, which should normally have supported them.

Fourth, 53% of the 2,537 destitutes surveyed on the pavements of Calcutta in September 1943 were female, and 64% of married destitutes were women. The pavement dwellers had earlier belonged to 820 families; half of these had broken up shortly before arrival, 70% due to husbands and wives separating. The women said they had left their villages for Calcutta because their husbands had either deserted them or asked them to search elsewhere for food (Greenough 1982: 220).

There is also anecdotal evidence of women being abandoned by husbands or being forced into beggary or prostitution, and of parents selling off girls (2–13 years of age) into prostitution and boys into bondage. The Nari Seva Sangh (a private women's service society), established in 1944 to shelter abandoned women during famine, reported: "women were being thrown out into the streets and multitudes of them were being forced into a life of shame" (quoted in Greenough 1982: 225).

This abandonment of women and children is not unique to the Great Bengal Famine. It is commonly observed during famines across different cultural contexts. Alamgir (1980: 135), for instance, notes that during the 1974 Bangladesh famine "there were many cases of desertion. In Rangpur, special homes for deserted children were set up. In Dacca, there were many women who were deserted by their husbands among the inmates of vagrant homes." Similarly, the women interviewed by Vaughan during the 1949 Malawi famine stressed "how frequently they were abandoned by men" (who had external social and economic linkages), and left with dying children on their hands (Vaughan 1987: 123).

I.3 Possible Explanations

Greenough and Appadurai offer cultural and sociobiological explanations for abandonment during the Great Bengal Famine. Greenough (1982: 215) argues that the

disintegration of families "did not occur randomly but seems to have been the result of the intentional exclusion of the less-valued family members [women and children] from domestic subsistence" by a decision of the male household head. He further argues that "determinations as to who will eat and who must starve are not made suddenly or arbitrarily in the heat of the famine but draw upon an existing cultural arrangement that assigns a value to family members according to their roles" (Geenough 1980: 208). Such decisions, he explains, are in line with the ideal of family continuity, equated with the continuity of the patriline through the adult male, whose survival thus counts over that of women and children: "If the separated 'master' survives a crisis, he is enabled to marry again despite the death or violation of his spouse, and he may then engender children who can be expected to maintain the lineage under more prosperous conditions" (Greenough 1982: 224). In this "explanation", the abandonment of the vulnerable is read as justified by the cultural notion of family continuity through males, rather than simply an act of individual self-interest by men who have a stronger fallback position. It also contradicts aspects of culture which value self-sacrifice and support of the poor and the disadvantaged. It is notable that despite their destitute state, women were found much more likely than men to have kept the children with them.

Appadurai (1984: 485) offers a somewhat different explanation. He suggests that the abandonment of women and children can be seen as "an effort to maximise the life-chance of each and every member of the family in circumstances where co-residence was clearly not feasible. Thus the sale of children . . . might be seen . . . as an effort to construct a better set of life-chances than those of the existing family structure." However, this explanation too is difficult to uphold in a social context where female chastity is highly regarded, and where the husband would know that the abandoned spouse would fall prey to sexual exploitation and so become a social outcaste. Abandoning wives and selling little daughters into prostitution can hardly be seen as a way of improving their life chances. Even if the woman managed to survive, she would not be accepted back socially either by her husband's family or by her birth family.[3] This is in fact what happened—in the aftermath of the famine, women filled the shelters opened to accommodate them.

An alternative way of understanding this process lies in recognizing the family as a unit of both cooperation and conflict, of both sharing and selfishness, where self-interest can also dominate, and outcomes can depend on differential bargaining power. Family breakup and female victimization can thus be seen as outcomes of shifts in relative male–female entitlements and fallback positions, and so in their relative bargaining strengths within the family. It is striking that the decision of the

[3] Greenough (1982: 225) recognizes this too; and this is also highlighted by research on the fate of women abducted during the partition of India (Menon and Bhasin 1998).

man to abandon his wife occurs at a point when the wife's entitlements appear to have collapsed completely while those of the husband are weakened but not entirely gone. It is a telling point that 57% of male but only 6% of female recipients of relief were found (in the survey mentioned earlier) to be able to pay cash for the food received, and also that the lowest excess mortality in absolute terms was among men in the age group 20–40. Moreover, one of the traditional (and few) fallback occupations for poor women in Bengal—paddy husking—was not sustaining. Paddy huskers ranked among the top four categories of those hardest hit by the famine (Sen 1981: 71); and the large number of new entrants to this occupation during the worst phase of the famine appears to have been symptomatic of women's growing marginalization in other occupations. The sudden congestion of this occupation would also have depressed the already low earnings of those who did find work. It is notable too that the abandoned women in the three illustrative accounts presented by Greenough all had husbands with resources to survive on. One account is of an 18-year-old found wandering in Calcutta, whose husband had pushed her out and "decided to stay on the [one acre of] land to look after his coming betel-vine". In another case, the husband had divorced and abandoned his wife to go join the army, and in the third account the husband had simply migrated out without informing her. In general, Greenough notes that the men stayed behind if they had resources in the village, or migrated to other areas if they had none—in either case abandoning the women and children while availing themselves of other options.

Men's fallback position was stronger than women's on several counts. First, men were more mobile and could migrate longer distances for jobs. Second, they were more likely to have some leftover assets, including land, while women's assets—typically small animals and jewelery—would by then have been disposed of, since it is common for families facing severe food crisis to sell off such movable assets first, and to hold on to large animals and land (Jodha 1975; Greenough 1982), which are typically owned by men. Support by friends and neighbors facing similar conditions also tends to collapse during famines, as happened here, and support from the state came *ex post facto*. In other words, women's work options and asset positions collapsed at a much earlier stage in the process of famine impoverishment than men's, leaving women with virtually no potential to contribute toward family survival and hence little bargaining power. Within the cooperative conflict framework and bargaining view of the family, at this point non-cooperation by the husband would make sense in the interests of his individual physical survival, and do him no harm in social terms (he could marry again).

It needs stressing that in these famine conditions the woman's fallback position diminishes *simultaneously* with her potential contribution to family income, since her lack of asset ownership and access to employment undermine both. Hence even though the husband's bargaining position improves as a result, it is still in his economic interest to abandon her, since an improved bargaining position would

provide him no *realizable* advantage, given the simultaneous (and severe) decline in her ability to contribute to joint well-being.[4]

Essentially, Greenough's view of the family (which Appadurai implicitly accepts) comes close to what Sen (1983) describes as the "despotic" family, where the male head takes all decisions and the others just obey, in line with Becker's unitary household model. What I have suggested here is that the process of family disintegration fits in better with the cooperative conflict (bargaining) view of the family as outlined by Sen (among others), even if he does not apply this view to his work on famines. In more recent theoretical and empirical developments on intra-household bargaining, it is this view of intra-family relations which is found to better explain unequal outcomes within the home than the unitary household model.[5]

At the same time, formal bargaining models of the household assume fully aware and typically self-interested individuals participating in the bargaining process. But what if women do not act in their own interest and therefore do not bargain to their best advantage? This brings me to the second aspect of my engagement with Sen's ideas, but more as contention than as extension.

II. False Perceptions or Structural Inequality?

Sen (1990) argues that the outcome of intra-household bargaining will be less favorable to a person the less value s/he attaches to her/his own well-being relative to the well-being of others ("perceived interest response"), and that this tends to be especially so in "traditional societies" such as India, where women tend not to think in terms of self-interest or their individual well-being:

> [I]nsofar as intrafamily divisions involve significant inequalities in the allotment of food, medical attention, health care, and the like (often unfavorable to the well-being—even survival—of women), the lack of perception of personal interest combined with a great concern for family welfare is, of course, just the kind of attitude that helps to sustain the

[4] Partha Dasgupta (1993: 329) misinterprets my 1990 paper in attributing to me the argument that a collapse of the woman's fallback position relative to her husband's *in itself* leads to her being abandoned in a famine. Equally, his argument that a weakening of the woman's fallback position, and hence the strengthening of the man's bargaining power, should encourage the man to remain in the marriage rather than abandon his wife, fails to take into account the fact that she has by then little left to contribute, and hence there is little for him to gain from "cooperation".

[5] See the references in Agarwal (1997). See also research undertaken at the International Food Policy Research Institute (IFPRI, Washington, DC; <www.ifpri.org>) which empirically tests different household models.

traditional inequalities. There is much evidence in history that acute inequalities often survive precisely by making allies out of the deprived. *The underdog comes to accept the legitimacy of the unequal order and becomes an implicit accomplice.*

(Sen 1990: 126; emphasis mine)

The idea that women tend not to have a clear perception of their individual interests in societies such as India—that is, that they may suffer from a form of "false consciousness", in effect making them complicit in perpetuating their unequal position—is interesting, but debatable. The empirical evidence that can be culled points more to the contrary.

For a start, it is difficult to infer from simply observing people's *overt* behavior whether they are conforming to an unequal order because they accept its legitimacy, or because they are afraid to challenge it, or because they lack exit options. To understand how women perceive unequal gender relations, we need to examine not only their overt acts of resistance (or lack thereof) but also the many covert ways in which they express their dissatisfaction. Empirical work which probes women's covert responses, by recording their views in contexts where they can express themselves freely, or by using participant observation methods to penetrate hidden expressions, provides diverse examples of women's "everyday resistance"[6] to intra-household inequalities in resource distribution and control, and to their double work burdens. There are, for instance, many recorded cases of South Asian rural women, despite being restricted by seclusion norms, covertly trying to get some cash which they can independently control, by taking up income-earning work secretly, or by clandestinely selling small amounts of household grain.[7] On the basis of their many interviews with village women in Bangladesh, Abdullah and Zeidenstein (1982: 47) observe:

> Women told us usually what *other* women have done. For example, one woman stocked rice in another woman's house so her husband could not know she had it. Another woman had a neighbour raise a goat for her so her husband would not know about it. Yet another woman has opened a *pan* [betel leaf] business with her young son and has told him to keep their earnings a secret from the husband. Most women say that they hide their savings in holes in the bamboo, in the roof, or under piles of cloth.

Similarly, Lindholm (1982: 201) notes that in Pakistan's North-West Frontier Province, "The husband considers this [i.e. a wife secretly selling grain] theft, but the wife considers it her just dues for her work." Risseeuw (1988: 278) finds that in Sri Lanka, women coir workers "usually hide their money in different parts of the house, so that, after a beating, [the woman] can disclose one place, thereby giving [the husband] the illusion she has handed all her savings to him".

[6] See Scott (1985) for an elaboration of this term.

[7] Abdullah and Zeidenstein (1982), Lindholm (1982), Luschinsky (1962), Nath (1984) and White (1992) give evidence from various parts of South Asia.

Most women spend the money so controlled on family subsistence, and especially on their children. However, again this need not imply false perceptions, since women's particular attention to children can indicate their distinct preferences[8] and/or their investment in social protection,[9] or old-age security (given their typical lack of other assets). Moreover, it is not uncommon for women to spend their hidden savings on personal needs (Abdullah and Zeidenstein 1982), or on building "social capital" by buying gifts for family members to win their support (Luschinsky 1962). Some invest in goats and cows which they keep in their parental homes as a form of social security, rather than sharing the earnings with spouses or in-laws (Nath 1984). Some others are found to circumvent unequal food-sharing in a joint family in ingenious ways, such as by holding clandestine picnics with women friends, as found in Nepal (Enslin 1990), or feigning spirit possession to extract food items otherwise denied them, as observed in Pakistan (Khan 1983) and north India (personal observation).

Interviews with peasant women by sociologists also establish that women do not accept the unequal gender division of labor as legitimate, whether they covertly resist it or merely lament it. White (1992: 318), for instance, observed that sometimes Bangladeshi village women served their husbands' friends tea without milk, so that the men would "not think she had nothing better to do than make tea for them all day and [would therefore] be discouraged from returning". Peasant women in north India comment:

> Agricultural labourer men help Jat men in the fields, but for Jat women it only means more work. We have to cook more food and feed the labourers as well. . . . Women should also have fixed hours of work.
>
> (Horowitz and Kishwar 1982: 17)

> We women stay at home and do back-breaking work even if we are feeling ill or if we are pregnant. There is no sick leave for us. But we do not have any money of our own and when the men come home we have to cast our eyes down and bow our heads.
>
> (Sharma 1980: 207)

Such examples challenge, in different ways, any simple notion that women in rural South Asia (or indeed elsewhere) have accepted the legitimacy of intra-household inequality.[10] The overt appearance of compliance ("cast our eyes down") need not mean that women lack a perception of their best interests; rather it can reflect a survival strategy stemming from the constraints on their ability to pursue those interests overtly (e.g. "we do not have any money of our own"). Hence

[8] See also Basu (2006) and Agarwal (1997) on situations where women may be willing to bargain with husbands on behalf of children, because they have preferences regarding children that are distinct from those of their spouses.

[9] See Agarwal (1997) on interest coalitions between mothers and children. Such coalitions with older children could also protect mothers from domestic violence.

[10] Women are also found to use covert forms of resistance in the workplace, especially if they have no formalized systems of expressing dissent (Gunawardena 1989).

although I agree with Sen (1990: 126) that it would be an error to take "the absence of protests and questioning of inequality as evidence of the absence of that inequality", I believe it would equally be an error to interpret the absence of *overt* protest as indicating that women accept the inequality. Complying with inequality need not imply complicity in its perpetuation.

Sen does recognize that deprived groups may comply for many different reasons—habit, hopelessness, resignation, etc.—but he attributes this to their acceptance of the legitimacy of the established order rather than their covertly resisting that order. He writes (Sen 1990: 127): "Deprived groups may be habituated to inequality, may be unaware of possibilities of social change, may be hopeless about upliftment of objective circumstances of misery, may be resigned to fate, and may well be willing to accept the legitimacy of the established order." My point of departure here is in emphasizing the many everyday but hidden ways in which women resist the established order.

It is likely, though, that this resistance varies by context. Hence on some issues women articulate or appear even to believe in ideologies that benefit men—for instance, maintaining that child care is women's responsibility; on other issues there is observable opposition, such as towards family authority structures, male control over cash, and domestic violence. Class factors might also affect the degree to which women see their self-interest as congruent with that of the household. In northern South Asia, it is middle-class and rich peasant women (who benefit from their husbands' properties and face greater social restrictions on outside employment), rather than women agricultural labourers, who typically insist that sons are important for continuing the lineage, and who have a more negative attitude toward daughters (Gardner 1990; Horowitz and Kishwar 1982; personal observation).

In order to explain gender inequalities I would therefore place much less emphasis than Sen does on women's incorrect perceptions of their self-interest, and much more on the constraints they face in pursuing that interest. The policy implications would thus be less toward making women realize (say, through awareness raising) that they *deserve* better, and more toward having them believe they can *do* better (by improving their self-confidence, their information access, etc.), and by helping them actually to do better, by strengthening their economic and social position.

I would also like to distinguish here between Sen's argument about women's false consciousness (which I have contested above) and his argument about *adaptive preferences*, with which I substantially agree, namely that the deprived may adapt their preferences according to what appears feasible. Sen, for instance, gives a striking example from a 1945 survey conducted a year after the Great Bengal Famine, which found that only 2.5% of the widows compared with 48.5% of widowers reported themselves "ill" or in "indifferent health", even though the former were widely recognized to be the more deprived (Sen 1984: 309). Sen notes (1999: 63): "The deprived people tend to come to terms with their deprivation because of the

sheer necessity of survival, and they may, as a result, lack the courage to demand any radical change, and may even adjust their desires and expectations to what they unambitiously see as feasible." I would add that this could happen not only if the deprived adapt their needs and preferences (as Sen and some others rightly emphasize), but equally if the deprived adapt what they *reveal* of their needs and preferences, according to who asks the questions and how, and what the respondent thinks the researcher wants to hear. Women may not reveal their real preferences overtly, but may do so through their covert actions.

III. Relative Capabilities, Domestic Violence and Women's Property Status

I now move to the third aspect on which I have engaged with Sen's ideas, namely capabilities and freedoms. This engagement represents both an extension and a contestation. It is an extension in that I move beyond the notion of absolute capabilities to that of relative capabilities. It is implicitly a contestation, in that I argue that higher absolute capabilities could in some contexts lead to perverse outcomes. It is also implicitly a contestation in two other respects: (1) I emphasize women's property status as a critical element in their capability enhancement and empowerment, an aspect that Sen has largely neglected while privileging women's employment status, and (2) I focus on hidden aspects of women's unfreedom in Kerala, such as a high incidence of domestic violence, as a counterpoint to Sen's overwhelmingly positive reading of women's position in that state.

Freedom is a key concept in Sen's definitions of capabilities and development. Sen defines capabilities as the freedom to choose what you have reason to value. And he defines development as a process of expanding the real freedoms that people enjoy. Development, he notes, "requires the removal of major sources of unfreedom" (Sen 1999: 3). Yet some serious forms of unfreedom have received rather little attention in discussions of capabilities or development. These are the unfreedoms hidden within families, in particular those stemming from domestic violence. If development means the expansion of human capabilities, then freedom from domestic violence should be integral to evaluating developmental progress. Actual evaluative exercises, such as the UNDP's Human Development Indices (for which Sen's capability approach provides the theoretical underpinnings), remain largely confined to conventional measures of well-being (e.g. income, education and longevity). Even the occasional broader formulation, such as in the Gender Empowerment Measure, neglects critical dimensions that can empower women, such as effective property rights and freedom from physical and mental abuse.

Although Sen himself does not grapple with this issue, many dimensions of his work are helpful in teasing out the adverse impact of domestic violence on such things as women's well-being, capabilities and agency, and on a country's overall development. As elaborated in Agarwal and Panda (2007), marital violence can undermine all the attributes on which Sen (1993, 1999) lays particular emphasis in terms of human development and well-being, such as a woman's capabilities and functionings (her economic, social and political freedoms and opportunities), and her well-being and agency goals (the goals she sets for her own advancement or for the advancement of others). By way of illustration, the serious physical and mental injury that can be caused by domestic violence can critically undermine a woman's economic freedoms—such as her capability of earning a living or acquiring property—by making her afraid of reprisal if she goes out to work, or seeks to upgrade her skills, or explores job options, or asserts her property rights. And when she suffers actual physical or mental injury it can affect her functioning in the job market by disrupting her work life, productivity and upward mobility. Domestic violence also undermines all five instrumental freedoms that Sen (1999) argues are central to evaluating human well-being and a country's development progress, namely protective security, economic facilities, social opportunities, political freedoms, and transparency guarantees.

At the same time, we also need to consider the reverse effect, namely the impact of given capabilities or functionings (or lack thereof) on the risk of domestic violence. Additionally, we need to look not just at the effects of a woman's *absolute* levels of capabilities or functionings on her risk of violence, but also at the effects of her *relative* capabilities or functionings *vis-à-vis* her spouse—an aspect that has received little attention.

III.1 Relative Capabilities and Perverse Outcomes

Normally we would expect that the greater a woman's educational level (functioning) and hence her job opportunities (capability), the higher her well-being. We would also expect any spousal *gap* in capabilities and functionings, which favors the woman, to improve her well-being outcomes by improving her bargaining power within the family. And we would expect her to be subject to less violence because her husband values her economic contribution. But consider a situation where a man married to a woman better employed than himself is irked by his wife's "superiority" and so beats her up. The injuries he causes can reduce her well-being outcomes both directly (e.g. by undermining her health and self-confidence) and indirectly (e.g. by reducing her earning abilities and social opportunities). In other words, in particular situations perverse effects can result, wherein a woman with higher absolute capabilities or functionings may be left worse off than one with

lower absolute capabilities or functionings, depending on her position *relative* to her husband's.

The extent of such perverse effects can, of course, vary with cultural context (including ideas about masculinity) and needs empirical testing. But irrespective of the empirical results, we need to recognize conceptually that well-being outcomes can depend not only on absolute capabilities and functionings but also on the play of relative capabilities and functionings, especially within families, and that this might sometimes lead to perverse effects.[11] In some respects, this recognition would help extend Sen's contribution on cooperative conflict within the family (Sen 1990), which he has linked rather little to his capability approach. At the same time, such a recognition would also complicate Sen's formulation, which does not deal with the possibility of perverse outcomes, such as husbands beating up wives who have higher capabilities (or functionings) than themselves.

Before illustrating this possibility empirically, an additional gap needs mention, namely, researchers' neglect of women's property status as a critical factor in women's well-being outcomes, sense of empowerment, and freedom from domestic violence.

III.2 Property Ownership versus Employment

The ownership of *immovable property* has several advantages over simply being employed. For a start, the security provided by property ownership does not vary with the vagaries of the labor market. A house or land also clearly indicates that the woman has a strong fallback position and tangible exit option.[12] This can deter violence. And if she still faces violence, owning (or otherwise having access to) a house or land can give her somewhere to go. A house, in particular, provides a ready roof over the head. But even with land she can build a homestead or set up a micro-enterprise, which brings in an income. Employment alone does not provide the same protection, for several reasons.[13] If women are simply unpaid workers on family farms or in family businesses, as is the case for many women in developing countries, employment *in itself* is unlikely to improve their financial situation. But even when women have earnings, these may be insufficient to rent a place if they need to escape violence. Also, for social reasons women may find it difficult to rent accommodation—in many cultures landlords are suspicious of single women

[11] In the voluminous literature on the capability approach, there is rather little discussion on how well-being outcomes can be affected by one person's capabilities relative to another's. Although Richardson (2007) and Iversen (2006) provide interesting takes on this, they do not focus on the potential for intra-household perverse outcomes.

[12] See also Agarwal (1994, 1997) on the potential impact of owning immovable property on bargaining power.

[13] Some studies find a lower incidence of violence among employed women, others a higher incidence, or no difference (see Agarwal and Panda 2007 for details).

tenants. A woman owning a home or land does not face the same problems. Moreover, in rural areas, land access enhances women's livelihood options and their overall sense of empowerment (Agarwal 1994). Property ownership can therefore reduce her risk of suffering violence by increasing her economic security, reducing her tolerance of violence, and providing a potential escape route should violence occur.

Would perverse capability effects also be less likely with property ownership compared to employment? We might indeed expect so. Employment alone, as noted above, typically does not provide a woman with an immediate, credible exit option. Hence a man wanting to show a wife who is better employed than himself her "proper place" within the relationship may not be deterred by the fact that she is employed. But owning a house or land visibly signals the strength of a woman's fallback position and clearly indicates that she has a tangible and credible exit route. Hence, a spousal gap in property ownership that favors the woman is less likely to lead to a perverse effect. What is important here is not whether a woman actually uses the exit option that immovable property provides, but that its very existence can deter the husband from abusing her.

III.3 Empirical Analysis

I jointly conducted (with Pradeep Panda) empirical analysis of the effect of women's capabilities (absolute and relative) on her risk of domestic violence, based on (1) his 2000/1 sample of 502 households (rural and urban) selected randomly from a population of ten wards in Trivandrum district in Kerala state (India), and (2) a panel resurvey of the same households by both of us in 2004/5 (which captured 80% of the earlier sample). The resurvey also provides information on the man's property status which the former did not (for details of both surveys see Agarwal and Panda 2007). The respondents were women aged 15 to 59 who had ever been married.

Table 9.1 gives the incidence of long-term physical violence (that is, where the women reported they had faced one or more incidents after marriage). This is cross-tabulated with the woman and her husband's individual and relative employment status and their individual and relative property status. We find that in and of itself, the woman's employment status does not protect her from violence. In fact there is a perverse effect—the incidence of physical violence is *higher* if a woman does seasonal or irregular work (as is usual with agricultural work) than if she is unemployed. Only where the woman has regular employment is the incidence of domestic violence lower than if she is unemployed (or does seasonal or irregular work). However, the gender gap in employment is again linked with a perverse effect. Where the woman has a higher employment status than her husband the

Table 9.1. Effects of absolute and relative capabilities on long-term physical violence against women

Capability	Number of cases	Long-term physical violence among cases (%)
	2000/1 survey (N = 502)	
Employment of woman respondent		
Unemployed	342	35.1
Regular employment	93	28.0
Seasonal or irregular employment	67	49.3
Employment of husband		
Unemployed	34	70.6
Regular	406	31.3
Seasonal or irregular	62	45.2
Spousal employment difference[a]		
Wife = husband	126	34.9
Wife < husband	360	34.4
Wife > husband	16	68.8
Immovable property owned by woman		
None	330	49.1
Land only	28	17.9
House only	71	9.9
House and land	73	6.8
Spousal immovable property ownership difference	*2004/5 resurvey (N = 402)*	
Neither spouse owns land or house	190	58.9 (112/190)
Woman is propertyless, husband owns land and/or house	68	23.5 (16/68)
Woman owns land and/or house, husband is propertyless	69	14.5 (10/69)
Both spouses own land and/or house	75	5.3 (4/75)

Note: W = wife, H = husband, u = unemployed, r = regular, s/i = seasonal or irregular.
[a] Spousal employment difference was calculated as follows:

(1) Wife = husband: wife and husband have similar type of employment {W(u), H(u); W(r), H(r); W(s/i), H(s/i)}
(2) Wife < husband: wife's employment status is lower than husband's {W(u), H(r); W(u), H(s/i); W(s/i), H(r)}
(3) Wife > husband: wife's employment status is higher than husband's {W(r), H(u); W(r), H(s/i); W(s/i), H(u)}

Sources: Adapted from Panda and Agarwal (2005: 835–6) and Agarwal and Panda (2007: 374).

incidence of violence is *twice* that of a woman who has the same or a lower employment status.[14] This empirically substantiates my point that it is important to take relative capabilities and functionings into account when assessing well-being outcomes.

[14] Duvvury and Allendorf (2001) similarly find greater domestic violence when the woman has a better employment status as, for instance, when the woman is employed and the man is unemployed.

Table 9.2. Women's property ownership, long-term physical violence, and their leaving and returning home (2000/1 survey)

Characteristics	All women		Propertyless women		Propertied women (owning land or house or both)	
	%	N	%	N	%	N
Among women facing long-term physical violence those who left home (N = 43)	24.0	43/179	19.1	31/162	70.6	12/17
Women who returned home among those who had left (N = 24)	55.8	24/43	67.7	21/31	25.0	3/12

Source: Adapted from Panda and Agarwal (2005: 838).

Second, unlike employment, women's property ownership is associated with a dramatically and *unambiguously* lower incidence of long-term physical violence. For instance, as many as 49% of the women who owned neither land nor house reported long-term physical violence. In contrast, the figure was 18% for those owning land, 10% for those owning a house, and 7% for those owning both.[15] Also, the preliminary results of the 2004/5 survey show that women are not negatively affected even when the gender gap in property ownership tilts in their favor. The incidence of violence is 14% if the woman owns property and the husband does not, but it is 24% if the husband owns property and the wife does not. In other words, owning property protects the woman, and there is no perverse relative capability effect, in that a propertied woman married to a propertyless man is *not* subjected to more violence.

Moreover, not only is the incidence of violence lower if a woman owns property, but she is also more likely to leave home and stay away if violence occurs, since she has somewhere to go. Of the 179 women experiencing long-term physical violence, 43 left home (Table 9.2). The percentage leaving home was much greater among the propertied (71%) than among the propertyless (19%). Also, of the 43 women who left home, although 24 returned, 87% of the returning women were propertyless. Few among the propertied women returned to their husbands. From the 2004/5 panel resurvey we also found that all the house-owning women who left home and did not return were living in their own houses, while 88% of the propertyless women who left home and did not return were living in their parents' home. These observations further support my contention that property ownership serves both

[15] Belonging to a matrilineal caste group (where property passes through the female line), however, does not make a difference over and above owning property (Panda and Agarwal 2005). Rather, the effect of matriliny is indirect, in that it enhances the chances of Kerala women (relative to women elsewhere) owning property, both among matrilineal communities and among others influenced by such communities.

as a deterrent to marital violence and as an exit option if violence does occur. As one woman said: "owning property is a powerful shield for women. It was because of these 10 cents of land [0.1 acre] and this small house [an outhouse] that I could escape from my in-laws' place when life became a nightmare."

The reduced risk of domestic violence to women if they own immovable property is further confirmed through logistic analysis, in which we controlled for the effect of a large number of other variables, such as the household's economic status, the educational and employment status of both spouses and any gender gap therein, the woman's age, duration of marriage, her access to social support, the woman and her husband's exposure to such violence in childhood, husband's alcohol abuse, and so on (see Agarwal and Panda 2007 for the results). The impact of a woman's property status was statistically significant whether she owned only land, or only a house, or both. The odds of being beaten if the woman owned both a house and land were found to be 20 times lower than if she owned neither. The odds were 11 times lower if she only owned a house, and eight times lower if she owned only land.

It is notable that the deterrent effects of property ownership are not specific to Kerala. Gupta (2006) found a similar link in West Bengal (east India): current physical violence was 38% among propertyless women and 15% among the propertied. Owning a house made a particular difference.

IV. The Neglected Issue of Women's Property Status

Women's property status forms a connecting thread through the diverse dimensions of intra-household gender relations on which I have focused (applying, extending or contesting aspects of Sen's wide-ranging work). It is an aspect which Sen has rarely touched on, beyond occasional references, but which I believe is of central importance in defining women's capabilities and functionings, freedoms and agency.

For instance, a critical factor in the collapse of women's fallback position during famine is the difference in the types of assets men and women typically own in South Asia. Women's assets, as noted, tend to be movables—small animals, jewelery, etc. They seldom own agricultural land or homesteads (except in Kerala and Meghalaya in India, where there are substantial pockets of traditionally matrilineal communities, and Sri Lanka, where most communities were either bilateral or matrilineal[16]). In Martha Chen's survey of rural widows in seven states, of the

[16] In matrilineal inheritance systems property passes through the female line; in bilateral systems it passes to and through both men and women.

470 widows whose fathers owned land only 13% inherited any of that land as daughters, Kerala being the exception, with 43% of the women inheriting (Agarwal 1998: 22). In Nepal, according to its 2001 census, only 17% of women own land or a house. Hence even within land-owning households, in extreme contexts such as famines, women and children can be abandoned and forced into destitution, since few typically own immovable assets. Owning property could protect women and children in such dire circumstances both directly and by reducing the chances of family breakup. This is also being recognized in other contexts of vulnerability, such as HIV/AIDS, as women who lack secure property rights face destitution, along with their children, when husbands die of the disease.[17] Similarly, as noted, women owning no immovable assets face significantly greater risk of domestic violence than those who own such assets, even in Kerala.

More generally, there is substantial evidence that if women own immovable property themselves, it makes a critical difference to their own and their children's economic and social well-being and overall empowerment. Owning even a small plot of land or a homestead enhances women's livelihood options and self-esteem, reduces their risk of poverty, and increases the likelihood of children surviving, attending school, and receiving health care (Agarwal 1994, 2003). With the growing feminization of agriculture in many parts of South Asia, land titles for women can also enhance agricultural productivity.

Indeed, it is no coincidence that the two regions in South Asia with relatively favourable human development indicators for women are Kerala in India and Sri Lanka as a whole. Neither have sex ratios adverse to women and hence "missing women" (Klasen and Wink 2006). Sen, in his interview in Agarwal *et al.* (2006: 353), concedes the importance of Kerala's matrilineal tradition (and hence openness to women owning property) as a factor in the state's favorable gender-related human development indicators, but in most of his writings he continues to emphasize female education and employment as the keys to women's advancement. I agree with him that we cannot depend simply on "historical luck" to improve women's position, and Kerala and Sri Lanka have certainly had historical advantages in women's property rights, but for those less favored by history, public policy could compensate by promoting land and housing rights for women, if policy-makers were persuaded to focus on this critically important, but sadly neglected, factor.

To conclude, there is much that Amartya Sen has written that we would agree with. There is some that he has written that we might disagree with. The important point, however, is not whether we agree with him on every aspect, but that we engage with the enormous wealth of ideas with which he has enriched us, and which challenge us in ways that the work of few thinkers of modern times can match.

[17] For evidence from South Asia, note the live testimonies presented by women at a UNDP Asia Pacific Court of Women on "HIV, Inheritance and Property Rights", Colombo, 18 August 2007 (see <www.sundaytimes.lk/070819/News/nws.18.html>).

References

Abdullah, T., and Zeidenstein, S. A. (1982), *Village Women of Bangladesh: Prospects for Change* (Oxford: Pergamon Press).

Agarwal, B. (1990), "Social Security and the Family: Coping with Seasonality and Calamity in Rural India", *Journal of Peasant Studies*, 17: 341–412.

—— (1994), *A Field of One's Own: Gender and Land Rights in South Asia* (Cambridge: Cambridge University Press).

—— (1997), " 'Bargaining' and Gender Relations: Within and Beyond the Household", *Feminist Economics*, 3(1): 1–51.

—— (1998), ' "Widows vs. Daughters or Widows as Daughters: Property, Land and Economic Security in Rural India", *Modern Asian Studies*, 32: 1–48.

—— (2003), "Gender and Land Rights Revisited: Exploring New Prospects via the State, Family and Market", *Journal of Agrarian Change*, 3(1–2): 184–224.

—— Humphries, J., and Robeyns, I. (eds) (2006), *Capabilities, Freedom and Equality: Amartya Sen's Work from a Gender Perspective* (Delhi: Oxford University Press).

—— and Panda, P. (2007), "Toward Freedom from Domestic Violence: The Neglected Obvious", *Journal of Human Development*, 8: 359–88.

Alamgir, M. (1980), *Famine in South Asia: Political Economy of Mass Starvation* (Cambridge, Mass.: Oelgeschlager and Hain Publishers).

Appadurai, A. (1984), "How Moral is South Asia's Economy?", *Journal of Asian Studies*, 43: 481–97.

Basu, K. (2006), "Gender and Say: A Model of Household Behaviour with Endogenously Determined Balance of Power", *Economic Journal*, 116: 558–80.

Becker, G. S. (1981), *A Treatise on the Family* (Cambridge, Mass.: Harvard University Press).

Chattopadhyay, K. P., and Mukerjee, R. K. (1946), *A Plan for Rehabilitation* (Calcutta: Statistical Publishing House).

Dasgupta, P. (1993), *An Enquiry into Well-Being and Destitution* (Oxford: Oxford University Press).

Dave, A., and Solanki, G. (2000), "Special Cell for Women and Children: A Research Study on Domestic Violence", in *A Summary Report of Four Records Studies* (Washington, DC: International Center for Research on Women), 25–33.

Duvvury, N., and Allendorf, K. (2001), "Domestic Violence in India: The Roles of Education and Employment", paper presented at the Sixth Women's Policy Research Conference on "The Status of Women: Facing the Facts, Forging the Future", Washington, DC, June 8–9.

Enslin, E. M. W. (1990), "The Dynamics of Gender, Class and Caste in a Women's Movement in Rural Nepal", unpubl. Ph.D. diss., Stanford University.

Gardner, K. (1990), "Paddy Fields and Jumbo Jets: Overseas Migration and Village Life in Sylhet District, Bangladesh", unpubl. Ph.D. diss., London School of Economics.

Greenough, P. R. (1982), *Prosperity and Misery in Modern Bengal: The Famine of 1943–44* (Oxford: Oxford University Press).

—— (1980), "Indian Famines and Peasant Victims: The Case of Bengal in 1943–44", *Modern Asian Studies*, 14: 205–35.

Gunawardena, A. N. (1989), "Bitter Harvest: Transformation of Women's Roles in the Sugar Economy of Sri Lanka", unpubl. Ph.D. diss., University of California at Los Angeles.

Gupta, J. (2006), "Property Ownership of Women as Protection for Domestic Violence: The West Bengal Experience", in ICRW (ed.), *Property Ownership and Inheritance*

Rights of Women for Social Protection: The South Asia Experience (Delhi: ICRW), 37–56.

Horowitz. B., and Kishwar, M. (1982), "Family Life: The Unequal Deal", *Manushi*, 1: 2–18.

Iversen, V. (2006), "Intra-Household Inequality: A Challenge for the Capability Approach?", in Agarwal *et al.* (2006), 106–31.

Jodha, N. S. (1975), "Effectiveness of Farmers' Adjustment to Risk", *Economic and Political Weekly*, 13(25): A37–A76.

Klasen, S., and Wink, C. (2006), "'Missing Women': Revisiting the Debate", in Agarwal *et al.* (2006), 286–327.

Lindholm, C. (1982), *Generosity and Jealousy: The Swat Pukhtun of Northern Pakistan* (New York: Columbia University Press).

Luschinsky, M. S. (1962), "The Life of Women in A Village of North India: A Study of Roles and Status", unpubl. Ph.D. diss., Cornell University.

Mahalanobis, P. C., Mukherjee, R., and Ghosh, A. (1946), *A Sample Survey of After-Effects of the Bengal Famine of 1943* (Calcutta: Statistical Publishing House).

Menon, R., and Bhasin, K. (1998), *Borders and Boundaries: Women in India's Partition* (New Brunswick, NJ: Rutgers University Press).

Minturn, L., and Hitchcock, J. T. (1966), *The Rajputs of Khalapur* (New York: John Wiley).

Nash, J. F. (1950), "The Bargaining Problem", *Econometrica*, 18: 155–62.

—— (1953), "Two-Person Cooperative Games", *Econometrica*, 21: 128–40.

Nath, J. N. (1984), "Dynamics of Socio-Economic Change and the Role and Status of Women in Natunpur: Case Study of a Bangladesh Village", unpubl. Ph.D. diss., University of Dhaka.

Panda, P., and Agarwal, B. (2005), "Marital Violence, Human Development and Women's Property Status in India", *World Development*, 33: 823–50.

Richardson, H. (2007), "Trust, Respect and the Constitution: The Social Background of Capabilities for Freedoms", *Journal of Human Development*, 8: 389–414.

Risseeuw, C. (1988), *The Fish Don't Talk About the Water: Gender Transformation, Power and Assistance Among Women in Sri Lanka* (Leiden: E. J. Brill).

Scott, J. C. (1985), *Weapons of the Weak: Everyday Forms of Peasant Resistance* (New Haven, Conn.: Yale University Press).

Sen, A. K. (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Delhi: Oxford University Press).

—— (1983), "Economics and the Family", *Asian Development Review*, 1: 14–26.

—— (1984), *Resources, Values and Development* (Oxford: Clarendon Press).

—— (1990), "Gender and Cooperative Conflicts", in I. Tinker (ed.), *Persistent Inequalities: Women and World Development* (New York: Oxford University Press), 123–49.

—— (1993), "Capability and Well-Being", in M. Nussbaum and A. K. Sen (eds), *The Quality of Life* (Oxford: Oxford University Press), 30–66.

—— (1999), *Development as Freedom* (Delhi: Oxford University Press).

—— (2006), "Continuing the Conversation", in Agarwal *et al.* (2006), 347–61.

Sharma, U. (1980), *Woman, Work and Property in North-West India* (London and New York: Tavistock Publications).

Vaughan, M. (1987), *The Story of an African Famine: Gender and Famine in Twentieth Century Malawi* (Cambridge: Cambridge University Press).

White, S. (1992), *Arguing with the Crocodile: Class and Gender in Rural Bangladesh* (London: Zed Books).

CHAPTER 10

FAMILY TIES, INCENTIVES AND DEVELOPMENT

A MODEL OF COERCED ALTRUISM

INGELA ALGER

JÖRGEN W. WEIBULL

I. Introduction

The strength of family ties appears to vary across countries and over time. In a recent paper, Alesina and Giuliano (2007) find that the strength of family ties has significant effects on various economic outcomes, such as labor market participation, the extent of home production and geographic mobility. Furthermore, it has been argued that individualism was an important force behind the Industrial

We are grateful to Daron Acemoglu, Philippe Aghion, Don Cox, Karen Norberg, Marcus Salomonsson and Yannick Viossat for helpful comments and discussions. We also thank audiences at Boston College, Boston University, Carleton University, HEC Montréal, SUNY Binghamton, Université de Cergy-Pontoise, Université Laval, the CEA, the EEA, and the Midwest Economic Theory meetings for feedback. The usual disclaimer applies. Jörgen W. Weibull thanks the Knut and Alice Wallenberg Research Foundation for financial support.

Revolution in England. For example, Max Weber thought that "the great achievement of . . . the ethical and ascetic sects of Protestantism was to shatter the fetters of the sib [the extended family]. These religions established . . . a common ethical way of life in opposition to the community of blood, even to a large extent in opposition to the family" (1951: 237). In his view, a strong sense of solidarity among members of the extended family, coupled with a hostile attitude towards strangers, promotes a culture where nepotism may thrive and counter the efficient development of markets. Likewise, Banfield (1958) thought that the "amoral familism" that he observed in certain parts of Italy was an impediment to economic development.

Motivated by the evidence that family ties vary in strength across cultures, we here pursue the line of thought suggested by Weber, by way of theoretically analyzing the effects of family ties on risk-sharing and incentives. We all face risk in the form of unknown realizations of future states of the world. Risk may in turn lead individuals to seek to pool resources and thereby mitigate severe adverse income shocks on the individual level. However, insurance markets are not well developed everywhere, and where markets are poorly developed the extended family tends to be an important source of insurance.[1] This observation leads to the following questions, which we seek to answer here. First, how does the insurance provided within the family affect work incentives for its members? Second, does such informal insurance represent an improvement compared to autarky, that is, each individual living by him- or herself without any access to insurance whatsoever? Third, can informal family insurance be sustained as a robust social norm in a society exposed to norm perturbations and migration opportunities? Fourth, can actuarially fair formal insurance, either by way of a compulsory and efficient social security system or by way of a perfectly competitive insurance market, outperform informal family insurance?

Since formal insurance pools the resources of a much larger number of individuals than informal family insurance does, it is to be expected that perfectly competitive market insurance would be more efficient than informal family insurance. However, while this certainly is true at a fixed level of individual risk, and with verifiable information about individuals' actions (here, work efforts), this reasoning neglects the fact that individuals' incentive to undertake risk-reducing efforts depends on the level of insurance and how the insurance is conditioned on their (verified) actions. Such moral hazard phenomena are analyzed in Helpman and Laffont (1975), Arnott and Stiglitz (1988, 1990) and, more recently, Bennardo

[1] In 2003 the total value of insurance premia (life and non-life) as a percentage of GDP was 12.48 in the US, 9.85 in France, 1.42 in Turkey and 1.74 in Mexico (OECD 2005). In their survey on private transfers between households, Cox and Jimenez (1990) conclude that in developing countries 20–90% of households receive (private) transfers, which can represent up to 20% of the average household income. In the US the corresponding figures are 15% and 1%, respectively. Since the average income of donor households exceeds that of recipient households (Cox, Galasso and Jimenez 2006), these transfers appear to provide some insurance. Using data from Thailand, Miller and Paulson (1999) show that remittances respond to shocks to regional rainfall.

and Chiappori (2003). Since formal insurance under such asymmetric information lowers the incentive for risk-reducing efforts, a competitive insurance market equilibrium can at best sustain partial insurance. While individuals' actions typically are their private information with respect to insurance companies, this is typically not the case, at least not to the same extent, with respect to family members who provide informal insurance to each other. Hence, one cannot rule out a priori the possibility that informal family insurance may outperform formal (market or public) insurance in some cases. One purpose of this chapter is to explore this possibility theoretically.

The family, on the other hand, may be particularly vulnerable to the Samaritan's dilemma (Buchanan 1975), which arises due to an altruist's inability to be commited to not helping out a person in need. This has been shown to lead to several inefficiencies, including suboptimal savings (Lindbeck and Weibull 1988) and underinsurance by poor individuals (Coate 1995). It further seems empirically relevant to study a closely related but distinct form of family tie, namely, the giving of transfers because of others' expectations rather than because of an "inner motive". Amartya Sen has long urged economists to seek a deeper understanding of the implications of the fact that people make many decisions within the realms of the family (see Sen 1983). As Sen put it: "The mixture of selfish and selfless behavior is one of the most important characteristics of group loyalty, [a mixture which] can be seen in ... kinship relations" (1987: 20). Such a tension between, on the one hand, the individual's selfish desires and, on the other hand, the surrounding society's expectations and social norms may be an important explanatory factor behind differences in economic growth and the different degrees of development in the world. An individual who lives in a society where he or she is expected to share his or her income with other family members, sometimes even with such relatively distant family members as first or second cousins,[2] may rationally expect to have to transfer so much of her income—if she is more successful than the others in the family—that the motive for making the effort in the first place is diluted. The same phenomenon occurs in partnerships between individuals who share output.

To study these questions we analyze a relatively simple model in which risk-averse individuals choose a costly risk-reducing action, "effort", that determines the probability distribution over output levels, as in Pauly (1974). We model the motive for sharing resources within a family as a mixture of voluntary and socially coerced altruism, or as a mixture of sympathy and commitment (as in Sen 1983).

[2] Dostie and Vencatachellum (2006) find that among 500 domestic workers in Tunisia, 40% see all their wages transferred directly to their fathers; among those who make voluntary transfers instead, the average amount remitted is 40% of their wages. According to Etounga-Manguelle (2000), people with a regular income in today's Africa are not only expected to provide help in emergency situations; they are also expected to finance the studies of younger members of the extended family and to contribute to the many lavish celebrations dictated by social rules; see also Mendelek Theimann (2006).

Voluntary altruism is formalized in the usual way, as a positive weight placed on others' welfare, while coerced altruism is formalized as a culturally or socially imposed norm for intra-family transfers. We determine the equilibrium behavior of individuals in this model, and use these results to seek to answer the questions we have posed. Moreover, we allow for the possibility that the answers to these questions may depend on the underlying riskiness of the environment, which in our model is captured by the ratio of the high to the low output level.

We first study atomistic individuals (with no family ties) living in autarky. The optimal autarky effort does not eliminate the risk of getting the low output. If one could pool the risks of a large number (technically a continuum) of individuals, a welfare-enhancing, budget-balanced insurance policy would thereby come into being. Insurance can only be second-best, however, due to moral hazard (an insurance policy cannot depend on the individual's effort). Since atomistic individuals do not anticipate being helped out by relatives, they would be willing to buy such an insurance policy: insurance markets are expected to develop in societies where family ties are so weak that no resource-sharing occurs within the family, independent of the underlying riskiness.

We then turn to the analysis of insurance provided by the family: formally we analyze this as a strategic interaction between two individuals who care about each other's material welfare, either voluntarily or under the coercion of social norms. An individual's total utility is taken to be a weighted sum of both individuals' material utility, which in turn is determined by each individual's work effort and consumption. The weight put on the other individual's material utility is assumed to be non-negative and not greater than the weight put on one's own material utility. This weight can be interpreted in terms of altruism, or, alternatively, in terms of the esteem derived from others who observe and evaluate one's behavior, such as members of one's extended family, village or society at large. By contrast to Alger and Weibull (2007b), where we assume that altruism alone determines transfers within the family, here we consider situations in which an individual's degree of altruism differs from that enforced by society. More precisely, we suppose that the interacting individuals live in a society with a social norm that dictates a larger transfer than the individuals' own altruism suggests. If the degree of such *coerced altruism* is strong, individuals feel forced to help each other out.[3]

The players invest effort in production, and output may be low or high. Once the outputs have been realized, these are observed by both individuals, and transfers occur. We assume that an individual is expected to make a transfer to his or her sibling only if his or her own output is high whereas the sibling's is low. We measure the level of coerced altruism by the share of the high output that a rich individual is expected to give to a poor sibling. Consumption is taken to equal the final

[3] Many individuals are willing to pay in order to avoid situations where they feel coerced to behave altruistically, even in the absence of potential social sanctions. For recent laboratory studies showing this, see Dana *et al.* (2006) and Broberg *et al.* (2007).

output available to the individual. In our model it does not matter whether or not individuals observe each other's efforts; this is because we focus on the case where utility is separable in effort and consumption.

In solving this two-player game we focus on the case of individuals with the same loglinear preferences over own consumption and effort.[4] This game has a unique Nash equilibrium. Its qualitative features are as follows. The anticipation of receiving a transfer when poor has a negative effect on an individual's incentive to exert effort. This free-rider effect is well known from other analyses of altruism. Hence, the equilibrium effort decreases as the degree of coerced altruism increases. By contrast, true altruism has a positive effect on an individual's incentive to exert effort: an altruist may exert more effort in order to have more to give to the other individual, an effect we call the "empathy effect" of altruism on effort.

Despite the previous strong emphasis in the literature on the possible moral hazard effect of intra-family altruism, there seems to be a limited number of empirical studies on this topic. Using data on farmer output in Mali, Azam and Gubert (2005) find that remittances from emigrated relatives have a negative impact on agricultural output. By contrast, Kohler and Hammel (2001) show, using census data for Slavonia from 1698, that the number of different crops grown by a family tended to *increase* as the nearby extended family increased. The authors were expecting the opposite effect, namely that as a result of insurance a family would invest less in risk-reducing planting strategies. However, our results suggest that there exists an intuitive explanation for this pattern: when a family expects to help another family out, the expected benefit of the risk-reducing planting strategy is increased.

We then find that while coerced altruism induces "involuntary" transfers *ex post*, such coercion is efficient *ex ante* in the sense that the equilibrium expected utility is higher than it would be in the absence of coercion—this is true even if individuals are selfish, so that no empathy effect arises. In such situations, it is as if social norms play the role of compulsory but informal insurance.

We use the equilibrium analysis to determine a sufficient condition for coerced altruism to be sustained as a social norm in a society with selfish individuals, where those who deviate from the norm suffer an exogenously administered punishment. Numerical simulations suggest that a more severe punishment would be required to sustain a given level of coerced altruism in societies where the underlying riskiness is high. This is quite intuitive, since the benefit from deviating from the norm by not transferring the socially sanctioned share of high output is higher when the difference between the outputs is large.

[4] This particular game has not been studied before. Most of the literature on altruism, starting with Becker (1974), assumes one-sided altruism (see also, for example, Bruce and Waldman 1990; Chami 1998; Lindbeck and Nyberg 2006). In models with two-sided altruism, typically only one of the players chooses an effort (see Laferrère and Wolff 2006 for a recent survey), or there is no risk (Lindbeck and Weibull 1988 analyze the effect of two-sided altruism on savings, and Chen and Woolley 2001 study the intra-household allocation of income on private and public goods).

Thanks to the mitigating effect of altruism on moral hazard, risk-pooling between individuals (who cannot observe each other's effort) may outperform a competitive insurance market. Compared to formal insurance among a large number of selfish individuals, the benefit of the empathy effect obtained with altruistically motivated transfers between few individuals (say, within a family) may outweigh the limitation inherent in pooling resources of few individuals, as compared with market insurance that pools the resources of a large number of individuals. We find that this occurs in environments with low underlying riskiness, since the mitigating effect of altruism on moral hazard is particularly pronounced in these environments. It is important to point out that, when comparing market and non-market insurance, we use the same measure of performance, i.e. for non-market insurance we do not take into account the welfare that the individuals derive simply from being altruistic.

Our baseline model is similar to that in Arnott and Stiglitz (1991): they also allow for an endogenous, risk-reducing effort, and they model family insurance as transfers within pairs of *ex ante* identical individuals. They address a different question, however: they ask whether, in the presence of insurance markets, supplemental informal insurance provided by the family improves welfare. Moreover, whereas in our model transfers within the family are driven by altruism, in their model family transfers are the outcome of a joint agreement. Thus, in their model, if the family members can observe each other's effort, the joint agreement specifies that total income should always be shared equally, and it specifies the *ex ante* optimal effort, conditional on equal sharing. Mathematically, this is equivalent to the special case in our model where siblings are fully altruistic towards each other (i.e. each sibling attaches the same utility weight to each other sibling's welfare as to his or her own). Arnott and Stiglitz also study the case where family members cannot observe each other's effort but can enforce transfer agreements: mathematically this is equivalent to the special case in our model where individuals are selfish (attach no weight to their sibling's welfare) and make transfers dictated by coerced altruism.

This chapter also complements a large literature that seeks to determine the potential of non-market insurance in the presence of constraints imposed by limited enforceability. Posner (1980) introduced the idea that non-market insurance could be feasible despite limited enforceability, and Kimball (1988) proved this formally. Coate and Ravallion (1993) characterized the best sustainable non-market insurance arrangement as a non-cooperative equilibrium in a repeated game between two self-interested individuals. Foster and Rosenzweig (2001) extended this analysis to allow for altruistic individuals, while Genicot and Ray (2003) determined properties of equilibria that deter not only deviations by single players but also by groups of players. Recently, Bramoullé and Kranton (2007) have shown that equilibrium risk-sharing networks where people form pairwise links generally provide partial insurance only. Following the empirical evidence that households are less than perfectly insured in many developing countries (see Townsend 1994, and the

literature that followed his initial contribution), Ligon, Thomas and Worrall (2002) tested the hypothesis that this is driven by limited enforceability. In contrast to this literature, in our model risk is endogenous through individuals' choice of effort, and we ask whether insurance markets can function if uninsured individuals expect to be helped out by their relatives if in need.

Our model is similar to that in Lindbeck and Nyberg (2006), who analyze altruistic parents' incentive to instill a work norm in their children. The incentive stems from parents' inability to commit themselves to not helping their children if in financial need. If the children feel a strong social norm to work (hard), then this reduces the risk that the children will be in need, which is good for the altruistic parents. On the other hand, the parents will suffer with the children if their work ethic is very demanding and the children fail. The parents instill just enough of the social work norm in their children so that these two effects are optimally traded off. While Lindbeck and Nyberg's model is asymmetric—parents are altruistic and move first and children are selfish—our model is symmetric—the two siblings move simultaneously and may be equally altruistic towards each other. Nevertheless, the issues dealt with are related, the models similar in structure, and the parametrization of preferences over consumption and effort identical.

The remainder of the chapter is organized as follows. In the next section we present and solve the basic decision problem of an atomistic individual. We then prove in section III that there exists a welfare-enhancing insurance policy. In section IV we introduce the family by presenting the basic game between two individuals, we show that this game has a unique equilibrium, and we conduct a comparative-statics analysis of the equilibrium outcome with respect to altruism and coerced altruism. We further analyze the sustainability of coerced altruism as a social norm. In section V we compare formal insurance with insurance provided within the family. Section VI concludes by summarizing our main results and by pointing to directions for future work. All of the mathematical proofs are available in Alger and Weibull (2007a).

II. Selfish Atomistic Individuals

We begin by presenting the backdrop, the same baseline model as in the companion paper Alger and Weibull (2007b). Consider a selfish individual who feels no wish or social pressure to help others, living in an environment where insurance is not available. The individual chooses an *effort level* that determines the probability distribution over the possible returns, or *output levels*. The output is either high, $y^H > 0$, or low, $y^L = y^H/\rho$, where $\rho > 1$, the ratio between the high and low output

levels, represents the *riskiness* of the physical environment; this is the fraction to which output is reduced in the "bad" outcome. We think of y^H as the *richness* of the environment.

We assume that an individual who makes an effort yielding a success probability $p < 1$ for the high output level and then consumes an amount $c > 0$ of the output achieves utility

$$u(c, p) = \ln c + \gamma \ln(1 - p), \tag{1}$$

where $\gamma > 0$. Hence, the expected utility associated with any choice of $p \in [0, 1]$ is

$$p \ln y^H + (1 - p) \ln y^L + \gamma \ln(1 - p) = \ln y^H - (1 - p) \ln \rho + \gamma \ln(1 - p).$$

As shown in Alger and Weibull (2007b), this utility specification corresponds to log-linear utility in consumption and effort, with γ as the marginal cost of *effort*, and with a success probability that is exponential in effort, tending to 1 as effort goes to plus infinity.[5] We will refer to the quantity in (1) as *material utility*. (For a selfish individual as here, utility *is* material utility, while for an altruistic individual, to be studied below, utility also includes others' material utility).

The success probability p^0 that maximizes expected utility is easily found to be

$$p^0 = \max \left\{ 0, 1 - \frac{\gamma}{\ln \rho} \right\}. \tag{2}$$

We note that the optimal effort level, when positive, is independent of the richness of the environment, y^H, is higher in a riskier environment, and is higher when the marginal cost γ of effort is lower.

Considering the possibility that the probability of success may be zero, the achieved expected material utility, or welfare, may be written:

$$u^0 = \max \left\{ \ln y^H - \ln \rho, \ln y^H + \gamma \ln \left(\frac{\gamma}{\ln \rho} \right) - \gamma \right\}. \tag{3}$$

Welfare is thus increasing in the richness of the environment, y^H, and decreasing in the riskiness of the environment, ρ, and in the marginal cost of effort, γ. The associated achieved expected output level, or *income*, is

$$y^0 = y^L + \left(y^H - y^L \right) p^0 = \max \left\{ \frac{1}{\rho}, 1 - \left(1 - \frac{1}{\rho} \right) \frac{\gamma}{\ln \rho} \right\} y^H.$$

Hence, the national income of an isolated country inhabited by such selfish individuals would be increasing in the richness of their environment and decreasing in their cost of effort. For low levels of riskiness, effort is not worthwhile at all, and then national income is decreasing in the riskiness, while for high levels of

[5] Let $x \geq 0$ be effort, let $u = \ln c - \beta x$, and let $p = 1 - e^{-\theta x}$, for $\beta, \theta > 0$. This is equivalent to (1) for $\gamma = \beta/\theta$.

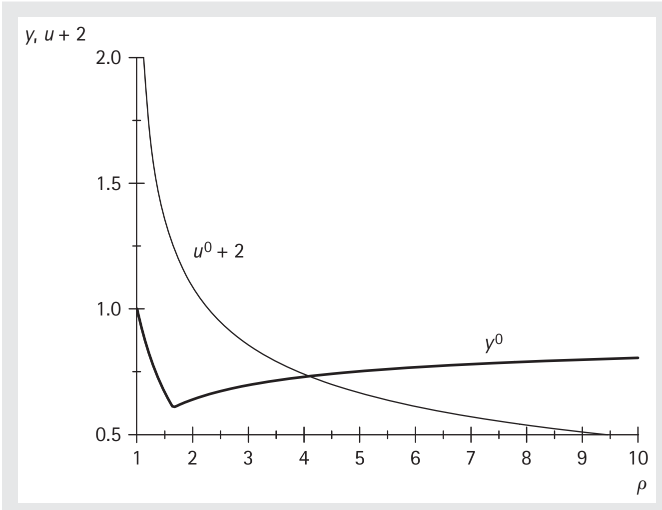


Fig. 10.1. Income (thick curve) and utility (thin curve) as a function of the riskiness ρ of the environment

riskiness, effort is worthwhile and national income is increasing in the riskiness of the environment. This non-monotonicity is illustrated in Figure 10.1.

III. Formal Insurance

Consider a large population of *ex ante* identical atomistic and selfish individuals facing i.i.d. risks of the type described above. By the law of large numbers, the fraction of individuals who end up with the low output is approximately $1 - p^0$, where p^0 is the optimal success probability (2) for each individual.

If neither individual efforts nor outputs (incomes) are verifiable, then the moral hazard problem for insurance systems is overwhelming, since the temptation will be strong for opportunistic individuals to insure themselves and then not exert any effort. A less overwhelming, but still significant, moral hazard problem pertains if individual efforts are not verifiable but outputs (incomes) are. We now analyze actuarially fair insurance in this case. Suppose that individual outputs (incomes) are verifiable while individual efforts are not. What would an optimal actuarially fair insurance give? If the individuals could collectively choose a compulsory insurance scheme—a form of social security—what would this be? What effect would insurance have on the incentive to exert effort and hence on national income?

We focus first on compulsory insurance policies. These may be provided either directly by a government agency or by perfectly competitive private insurance companies. Formally, an insurance policy is a pair $(\sigma, \pi) \in [0, 1]^2$, where the insurance premium is πy^H and the indemnity, net of the premium, to an individual who obtains the low output is σy^H. Thus, an individual who receives the high output enjoys consumption $(1 - \pi)y^H$ while an individual who receives the low output enjoys consumption $(\rho\sigma + 1)y^H/\rho$. Given such an insurance policy (σ, π), the individually optimal effort level is the same for all individuals. The decision problem faced by each individual is equivalent to that in autarky, with the original riskiness ρ replaced by the (lower) riskiness

$$\hat{\rho} = \frac{1 - \pi}{\rho\sigma + 1}\rho. \tag{4}$$

Hence, from (2) we obtain the optimal success probability,

$$\hat{p} = \max\left\{0, 1 - \frac{\gamma}{\ln \hat{\rho}}\right\}. \tag{5}$$

Under insurance (σ, π), each individual lives in a less risky environment and thus exerts less effort than in autarky. For $\gamma < \ln \hat{\rho}$ all individuals exert positive effort under the insurance policy. For $\ln \hat{\rho} < \gamma < \ln \rho$, no individual exerts any effort under the insurance policy, although all individuals would have exerted effort in autarky.

We define an insurance policy $(\sigma, \pi) \in [0, 1]^2$ to be *optimal* if it maximizes individual expected utility under the constraint that the policy is actuarially fair: $\hat{p}\pi = (1 - \hat{p})\sigma$. In order to characterize the optimal insurance policy for given values of ρ and γ, one should view the individually optimal success probability $\hat{p}$, defined in equation (5), as a function of the insurance policy (σ, π) and express each individual's achieved expected utility as

$$\hat{u}(\sigma, \pi) = \hat{p}(\sigma, \pi)\ln(1 - \pi) + [1 - \hat{p}(\sigma, \pi)]\ln(\sigma + 1/\rho) + \gamma \ln[1 - \hat{p}(\sigma, \pi)].$$

This defines $\hat{u} : [0, 1]^2 \to \mathbb{R}$ as a continuous function of σ and π. By Weierstrass's Maximum Theorem there exists an optimal insurance policy. It is not difficult to show that in an environment in which effort is worthwhile in autarky, insurance is welfare-enhancing:

Proposition 1. *There exists at least one optimal insurance policy* $(\hat{\sigma}, \hat{\pi}) \in [0, 1]^2$. *The no-insurance policy,* $\sigma = \pi = 0$, *is suboptimal if* $\gamma < \ln \rho$.

Until now we have assumed that insurance is compulsory. Suppose now that it is voluntary. If it is provided by the government, and each individual can choose whether or not to buy one insurance policy $(\hat{\sigma}, \hat{\pi})$, then all individuals will do so, according to Proposition 1. Suppose instead that insurance is voluntary and provided by private insurance companies. In the face of the moral hazard problem—the

non-verifiability of individuals' efforts—can the optimal policy be obtained by way of a perfectly competitive market? Assume, then, that there are risk-neutral profit-maximizing insurance companies who cannot verify individual efforts, only outputs (incomes). Assume also that insurance is indivisible, that is, individuals can buy either one or no insurance policies. If all insurance companies were to offer the optimal insurance policy $(\hat{\sigma}, \hat{\pi})$ characterized above, then it follows from Proposition 1 that each individual would buy the policy (and insurance companies would just break even). There would be no way for an individual insurance company to earn positive profits by way of offering another policy (σ, π). In sum: if multiple or partial purchases of insurance policies are not possible, then a perfectly competitive insurance market can deliver the optimal insurance policy to a population of atomistic and selfish individuals.

While Proposition 1 establishes that each individual's achieved expected utility is higher under optimal insurance than in autarky, expected output, and hence national income, is lower:

$$\hat{y} = y^L + (y^H - y^L)\hat{p} < y^L + (y^H - y^L)p^0 = y^0.$$

Formal insurance markets tend to be well developed only in some countries. In many places individuals have little or no access to formal insurance, and are instead involved in risk-sharing arrangements at the family or the village level. We turn next to the analysis of a society where such risk-sharing is triggered by an expectation that individuals help out less fortunate family members.

IV. Coerced Family Altruism

Following in the tradition of Smith (1790) and Edgeworth (1881) among other economists, Sen argues that individuals are not solely driven by a wish to maximize their own material welfare. In particular, Sen identifies "sympathy" and "commitment" as two other motives: "when a person's sense of well-being is psychologically dependent on someone else's welfare, it is a case of sympathy", and "one way of defining commitment is in terms of a person choosing an act that he believes will yield a lower level of personal welfare to him than an alternative that is also available to him" (Sen 1977: 327). In this section we inject both some sympathy and some commitment into our model. Thus, assume now that individuals still work individually but belong to families, and that in case individual output levels between siblings are unequal, a rich sibling is expected to transfer resources to a poor sibling. Siblings may feel sympathy towards each other, and hence a wish to help each other, but we focus here on the case of a social norm dictating transfers

that are at least as large as the transfers that would be given voluntarily in the absence of the social norm. In other words, people may feel coerced to behave as if they were more altruistic towards other family members than they actually are. Throughout section IV we take such a social norm as given. In section IV.3 we analyze the robustness of such social norms.

More precisely, assume that each individual i has one sibling, i', and each such pair interacts over two periods, along the lines of the model in the preceding section. Thus, in the first period, both siblings simultaneously choose their success probabilities p_i and $p_{i'}$. The output y_i of individual i is realized at the end of the first period. For the sake of notational and analytical convenience, we take the two siblings' outputs to be statistically independent random variables. This independence assumption can be relaxed to allow for positive correlation between outputs, but this does not give more insight into the incentive problems analyzed here (see Alger and Weibull 2007b).

At the beginning of the second period, both siblings observe each others' outputs, and each sibling makes the socially sanctioned transfer to his or her sibling. After these transfers, each individual's *disposable income*, or *consumption*, therefore equals his output plus any transfer received from the sibling minus any transfer given to the sibling. Since only the net transfer matters for the final utility, it suffices to define the *socially sanctioned net transfer* from individual i to his or her sibling i' when the output vector is $(y_i, y_{i'})$, which we denote $\tau_i(y_i, y_{i'})$. We assume that an individual who gets a high output is expected to make a net transfer $T \in [0, 1]$ to his or her sibling should the latter's output be low. Thus, $\tau_i(y_i, y_{i'}) = T$ if $(y_i, y_{i'}) = (y^H, y^L)$, $\tau_i(y_i, y_{i'}) = -T$ if $(y_i, y_{i'}) = (y^L, y^H)$, and $\tau_i(y_i, y_{i'}) = 0$ if $y_i = y_{i'}$.

We analyze this interaction as a game, denoted G, in which a pure strategy for individual i is a success probability $p_i \in [0, 1]$. Each strategy profile determines the *total utility* to sibling i in each state $(y_i, y_{i'})$:

$$\ln(y_i - \tau_i(y_i, y_{i'})) + \alpha \ln(y_{i'} + \tau_i(y_i, y_{i'})) + \gamma \ln(1 - p_i) + \alpha\gamma \ln(1 - p_{i'}),$$

where $\alpha \geq 0$ represents the degree of *true mutual altruism* between the siblings.

We assume that the socially coerced transfer from rich to poor siblings takes the form $T = \tilde{t} y^H$, where

$$\tilde{t} = \frac{\rho\tilde{\alpha} - 1}{\rho\tilde{\alpha} + \rho} \tag{6}$$

and $\tilde{\alpha} \in [1/\rho, 1]$. This form is analytically convenient, but also natural, since $\tilde{t} y^H$ is the transfer that a rich individual with true altruism $\tilde{\alpha}$ would choose to make voluntarily to a poor sibling (see Alger and Weibull 2007b). Hence, our assumption that the social norm dictates a transfer that is not smaller than what siblings would freely choose may be succinctly expressed as $\tilde{\alpha} \geq \alpha$.[6]

[6] By contrast, Alger and Weibull (2007b) assume that transfers are voluntary and driven solely by true altruism. This corresponds to the special case $\tilde{\alpha} = \alpha$.

IV.1 Equilibrium and Comparative Statics

In the first period, each individual independently chooses a success probability. The (*ex ante*) *expected total utility* for individual i, which is also his or her payoff in game G, can be expressed as a function of the two success probabilities:

$$\begin{aligned} U_i(p_i, p_{i'}) = {} & (1+\alpha)\ln y^H - (1-p_i)(1-p_{i'})(1+\alpha)\ln\rho \\ & + p_i(1-p_{i'})[\ln(1-\tilde{t}) + \alpha\ln(1/\rho+\tilde{t})] \\ & + p_{i'}(1-p_i)[\ln(1/\rho+\tilde{t}) + \alpha\ln(1-\tilde{t})] \\ & + \gamma\ln(1-p_i) + \alpha\gamma\ln(1-p_{i'}). \end{aligned} \tag{7}$$

Each player has a unique best reply to the other's strategy. The following proposition characterizes the unique Nash equilibrium of G.

Proposition 2. *Suppose that the two siblings are* ex ante *identical, with a common degree of altruism* α, *and live in a society with coerced altruism* $\tilde{\alpha} \geq \max\{1/\rho, \alpha\}$. *Then* G *has a unique Nash equilibrium. This equilibrium is symmetric, with common effort zero if*

$$\ln\left(\frac{\rho+1}{\rho+\tilde{\alpha}\rho}\right) + \alpha\ln\left(\frac{\tilde{\alpha}+\tilde{\alpha}\rho}{\tilde{\alpha}+1}\right) + \ln\rho \leq \gamma.$$

Otherwise the common success probability $\tilde{p}$ *is the unique zero in* $(0,1)$ *of*

$$(1-p-\alpha p)\ln\left(\frac{\rho+1}{\rho+\tilde{\alpha}\rho}\right) + (\alpha-p-\alpha p)\ln\left(\frac{\tilde{\alpha}+\tilde{\alpha}\rho}{\tilde{\alpha}+1}\right) + \ln\rho = \frac{\gamma}{1-p}.$$

Note that, even if $\ln\rho > \gamma$ so that the autarky effort p^0 is positive, here the success probability may be nil if $\ln\rho > \gamma$. To a relatively selfish individual, the prospect of having to support a poor sibling does not give much incentive to work hard and the prospect of being helped out if poor has a disincentive effect. In fact it is not difficult to show that α, the common degree of true altruism, has a positive incentive effect on the success probability while the degree of coerced altruism, $\tilde{\alpha}$, has a negative effect:

Proposition 3. *Suppose that the two siblings are* ex ante *identical, are truly altruistic towards each other of degree* $\alpha \in [0,1]$, *and live in a society with coerced altruism* $\tilde{\alpha} \geq \alpha$, *where* $\rho\tilde{\alpha} > 1$. *Their equilibrium success probability* $\tilde{p}$ *is increasing in* α *and decreasing in* $\tilde{\alpha}$.

Increasing true altruism entails an *empathy effect*: the individuals take into account the utility of the other individual to a larger extent, and therefore have an incentive to increase their effort. Increasing coerced altruism entails a *free-rider effect*: an increase in $\tilde{\alpha}$ means that the individual gets to keep less should he be lucky and his sibling unlucky, and gets a larger transfer in the opposite case.

IV.2 Welfare

To make comparisons with settings where individuals are selfish we use the equilibrium-expected material utility $\tilde{u}$ as a welfare measure. Proposition 2 defines the equilibrium success probability as a function of both true altruism α and coerced altruism $\tilde{\alpha}$, so we may write:

$$\begin{aligned}\tilde{u}(\alpha, \tilde{\alpha}) = \ln y^H + \tilde{p}(\alpha, \tilde{\alpha})[1 - \tilde{p}(\alpha, \tilde{\alpha})] \ln\left[\left(1 - \tilde{t}\right)\left(1/\rho + \tilde{t}\right)\right] \\ - [1 - \tilde{p}(\alpha, \tilde{\alpha})]^2 \ln \rho + \gamma \ln [1 - \tilde{p}(\alpha, \tilde{\alpha})]. \qquad (8)\end{aligned}$$

We find that $\tilde{u}$ is increasing in the common degree of true altruism, for any given degree of coerced altruism, *ceteris paribus*:

Proposition 4. *For any $\rho > 1$, $\gamma > 0$ and $\tilde{\alpha} > 1/\rho$, $\tilde{u}$ is increasing in α, for $\alpha \leq \tilde{\alpha}$.*

In our companion paper we analyze the special case $\tilde{\alpha} = \alpha$, i.e. transfers that are driven solely by true altruism. There we show that the level of true altruism that maximizes the expected material utility is full altruism, $\alpha = 1$. Together with Proposition 4 this result implies that in our model the vector of coerced and true altruism $(\tilde{\alpha}, \alpha)$ that maximizes the expected material utility involves sharing output equally, $\tilde{\alpha} = 1$, and full true altruism, $\alpha = 1$.

Coerced altruism arguably exists in many parts of the world. Can such a social or cultural norm be beneficial in terms of (*ex ante*) expected material utility? In order to analyze this question in its starkest form, consider a pair of completely selfish siblings, $\alpha = 0$, who live in a society with coerced altruism of degree $\tilde{\alpha} > 1/\rho$. Equation (7) gives an expression for the *ex ante* expected total utility in equilibrium. In the special case of selfish siblings, this is the same as the *ex ante* expected material utility $\tilde{u}(0, \tilde{\alpha})$, where $\tilde{u}(0, \tilde{\alpha})$ is defined in equation (8). Figure 10.2 plots this expected material utility against the level $\tilde{\alpha}$ of coerced altruism, for $\rho = 5$ and $\gamma = 0.5$.

As $\tilde{\alpha}$ increases beyond $1/\rho$, the smallest value of $\tilde{\alpha}$ for which a positive transfer may occur, we see that the expected material utility increases. Hence, here the marginal "benefit" of coerced altruism (the imposed mutual insurance) outweighs its marginal "cost" (reduced effort). In this example, coerced family altruism is thus advantageous *ex ante*, in spite of its adverse effect on effort. This can be shown more generally: selfish siblings are better off in a society with coerced altruism than living in autarky.[7] Moreover, the optimal level of coerced altruism does not dictate that siblings should split evenly—the rich sibling should keep more than half of the joint output:

[7] In a similar vein, Arnott and Stiglitz (1991) show that a pair of selfish siblings who cannot observe each other's efforts (and therefore cannot enforce transfers conditional on efforts) would nonetheless gain by agreeing *ex ante* to a contract whereby a rich sibling would make a transfer to a poor one.

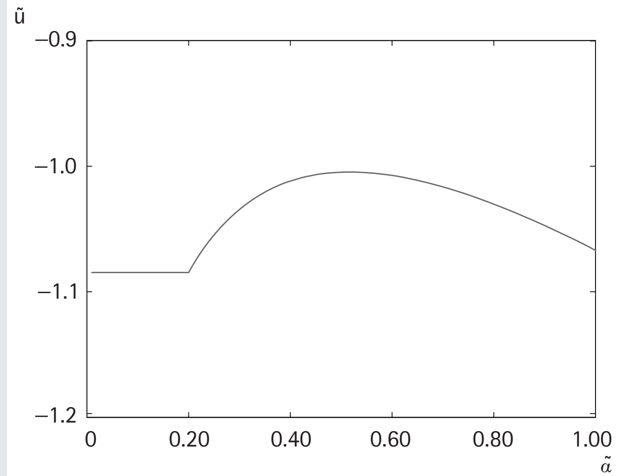


Fig. 10.2. Expected material utility ũ $(0, \tilde{\alpha})$ of selfish individuals with coerced altruism $\tilde{\alpha}$ (for $\rho = 5$ and $\gamma = 0.5$)

Proposition 5. *In a society with selfish individuals* $(\alpha = 0)$, *the level of coerced altruism that maximizes their expected material utility exceeds* $1/\rho$ *and is less than 1*:

$$\frac{1}{\rho} < \arg \max_{\tilde{\alpha} \in [0,1]} \tilde{u}(0, \tilde{\alpha}) \leq 1.$$

The intuition for the optimal $\tilde{\alpha}$ to exceed $1/\rho > 0$ is as follows. First, for $\tilde{\alpha} < 1/\rho$ no transfers are given and thus efforts are constant when $\tilde{\alpha}$ is changed. Consequently the expected material utility is also constant. At $\tilde{\alpha} = 1/\rho$, however, material utility increases when $\tilde{\alpha}$ is marginally increased. An increase in $\tilde{\alpha}$ will obviously increase the transfer between a rich and a poor sibling. Because of the concavity of the utility from consumption, this increased insurance will enhance both siblings' *ex ante* expected material utility *if they would not change their efforts*. Consider now the effort choice by one of the siblings, i. By the Envelope Theorem, a marginal increase in $\tilde{\alpha}$ will have no first-order effect on i's expected utility. However, the other sibling, i', will also change his or her effort. But it turns out that precisely at $\tilde{\alpha} = 1/\rho$, this marginal change in behavior also has no first-order effect on the expected material utility of sibling i.

While the expected utility to selfish siblings is higher under coerced altruism than without it, effort and thus the expected output $\tilde{y}$ is lower, where

$$\tilde{y} = y^L + (y^H - y^L)\tilde{p} < y^L + (y^H - y^L)p^0.$$

IV.3 Coerced Altruism as a Robust Social Norm

Which degree of coerced family altruism will prevail as a social norm in the long run, in a population consisting of identical selfish individuals, in a given physical environment? Suppose that a certain degree of coerced altruism, $\tilde{\alpha}$, is a social norm in a subset of the population, group A, while another degree of coerced altruism, $\tilde{\alpha}'$, is the social norm in another subset of the population, group B. Suppose, moreover, that each group is in its corresponding equilibrium, as defined in section IV.1. Thus $\tilde{u}(0, \tilde{\alpha})$ is the expected material utility in group A and $\tilde{u}(0, \tilde{\alpha}')$ that in group B (see equation (8)). A situation in which $\tilde{u}(0, \tilde{\alpha}') < \tilde{u}(0, \tilde{\alpha})$ does not appear to be socially stable. All individuals being selfish by assumption, individuals, families and local communities in group B would be keen to adopt the social norm of group A, since that results in higher welfare. Such a change of social norms could occur either by adaptation or by migration. Given such destabilizing possibilities, we call a degree of coerced altruism $\tilde{\alpha}$ *socially robust* if there exists no degree of coerced altruism $\tilde{\alpha}' \neq \tilde{\alpha}$ such that $\tilde{u}(0, \tilde{\alpha}') > \tilde{u}(0, \tilde{\alpha})$. Clearly a socially robust degree of altruism, according to this definition, is an optimal degree identified in Proposition 5.

Another relevant question is the enforcement of a given degree of coerced altruism. We will not go into any depth here, making only one observation. Suppose that coerced altruism of some degree $\tilde{\alpha}$ is an internalized social norm in a population. Individuals who give a smaller transfer than $\tilde{t}$ to their sibling, when that sibling is poor and they themselves are rich, experience disutility, and the more disutility the larger the fraction of the population (or their peer group) that adheres to the social norm. Formally, this amounts to a modification of the model above as follows. Consider a pair of identical siblings, both selfish. Let $\tilde{p}$ be the equilibrium success probability (defined in Proposition 2) and let $z \in [0, 1]$ be the share of the population who adhere to the norm by way of giving a transfer $t \geq \tilde{t}$ when rich and their sibling is poor (where $\tilde{t}$ is defined in equation (6)). Let $\varphi(z)$ be the disutility experienced by an individual who deviates from this norm, where φ is an increasing function with $\varphi(0) = 0$ and $\varphi(1) = \lambda > 0$. Since by assumption all individuals are selfish, a deviator will not give any transfer to his or her sibling. Hence, a deviator will make an effort such that the success probability p solves

$$\max_{p\in[0,1]} \ln y^H - (1-p)(1-\tilde{p})\ln\rho - p(1-\tilde{p})\varphi(z) + (1-p)\tilde{p}\ln(1/\rho+\tilde{t}) + \gamma\ln(1-p). \quad (9)$$

Let p^d be the (generically unique) solution to this program.

In this generalized model of coerced altruism among identical selfish individuals, coerced altruism of degree $\tilde{\alpha}$ will thus prevail as a social norm, $z = 1$, if and only if

$$\lambda p^d(1-\tilde{p}) + \tilde{p}(1-\tilde{p})\ln\left[(1-\tilde{t})\left(1/\rho+\tilde{t}\right)\right] - (1-\tilde{p})^2\ln\rho + \gamma\ln(1-\tilde{p}) \geq -(1-p^d)(1-\tilde{p})\ln\rho + (1-p^d)\tilde{p}\ln(1/\rho+\tilde{t}) + \gamma\ln(1-p^d). \quad (10)$$

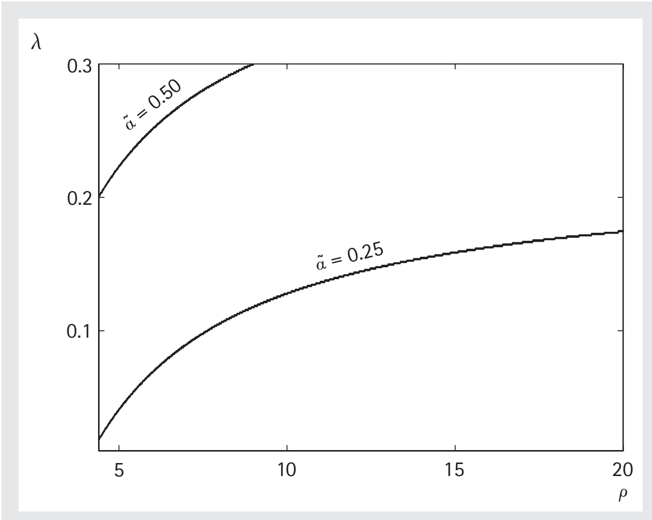


Fig. 10.3. Sustainability of coerced altruism as a social norm, as a function of ρ and λ

We also note that, viewed as a game with a continuum of players, there may exist multiple equilibria for certain parameter combinations; it is possible that both $z = 1$ and $z = 0$ are equilibria.[8] In the first equilibrium, all individuals adhere to the social norm of coerced altruism of degree $\tilde{\alpha}$ while in the second equilibrium nobody gives any transfer (and experiences no disutility from doing so, since everybody deviates).

Is coerced altruism easier to sustain as a social norm in a harsher climate? We can illustrate this by studying the range of λ-values for which the adherence condition (10) is met, as a function of ρ. Thus, in Figure 10.3 the lower curve shows the lowest value λ, of the cost inflicted on deviators, for which coerced altruism $\tilde{\alpha} = 0.25$ may be sustained as a social norm (for $\gamma = 0.5$). As expected, the temptation to deviate from the norm is larger for larger values of ρ: the threshold value increases as ρ increases. The upper curve in the figure further shows that, as expected, for any ρ a larger punishment would be required to sustain a higher level of coerced altruism, $\tilde{\alpha} = 0.5$ (see the upper curve).

In brief, our analysis shows that informal insurance by way of coerced altruism is beneficial, and that it may arise as a social norm. However, intuition would suggest that formal insurance institutions, which pool the risks of large numbers of individuals, would be even better. It is indeed easy to establish that selfish individuals would rather live in a society where they have access to formal insurance than in one with coerced altruism, even if coerced altruism is at its optimal level:

[8] Indeed, there may also exist a whole continuum of intermediate equilibria where the share of population z is such that all individuals are indifferent between adhering to the norm or deviating from it.

Proposition 6. *For selfish individuals, the expected material utility under coerced family altruism is lower than under optimal compulsory insurance.*

However, this does not answer the question whether coerced altruism, if coupled with some true altruism, would also be dominated by formal insurance. We analyze this question next. In particular, we analyze incentives for people to migrate from developing to developed countries, and compare the welfare of immigrants to that of the local population.

V. Formal Insurance versus Coerced Altruism

For the purposes of this thought experiment we assume that coerced altruism is non-binding, i.e. $\alpha = \tilde{\alpha}$, and we will refer to α as the strength of family ties. By virtue of Proposition 4 this assumption gives informal insurance the best possible chance against formal insurance.

Consider a pair of siblings living in some country, A, where no formal insurance exists, but where family ties are very strong, $\alpha_A = 1$. Assuming that this sibling pair would keep its strong family ties if it were to move to another country, would it have an incentive to migrate to country B, where family ties are weak, $\alpha_B = 0$, and where an optimal compulsory insurance system (for the inhabitants of B) is in place? If so, how would this pair of siblings fare compared to the original inhabitants of country B, in terms of material utility and income? Let (y_A^H, ρ_A) and (y_B^H, ρ_B) denote the sibling pair's environment in countries A and B, respectively. An incentive for the sibling pair to migrate would obviously exist if $y_B^H > y_A^H$ and $\rho_A < \rho_B$; then the expected material utility, and hence also total (altruistic) utility, would be higher in country B. Consider, therefore, a situation in which the siblings' incomes, when successful, are the same in both countries, $y_A^H = y_B^H = 1$, but the riskiness may differ (and country B has an insurance scheme).

In country B the compulsory insurance policy $(\hat{\sigma}_B, \hat{\pi}_B)$ entails an income riskiness, $\hat{\rho}_B$, derived from the triplet $(\rho_B, \hat{\sigma}_B, \hat{\pi}_B)$ according to equation (4). If the pair of siblings moved from country A to country B, they would represent a negligible part of the population there, and would therefore have no effect on the optimal insurance contract. They may therefore view the income riskiness $\hat{\rho}_B$ in country B as fixed and given. Since the optimal insurance policy $(\hat{\sigma}_B, \hat{\pi}_B)$ is second best, $\hat{\rho}_B > 1 = \alpha_A$, a successful immigrant sibling (the one obtaining the high output) would make a transfer to his or her unsuccessful sibling (the one obtaining the low output) even in country B. As long as $\hat{\rho}_B \leq \rho_A$, the siblings would be as well off as

in their country of origin. In fact, the next proposition implies that if this inequality is strict, they would be strictly better off in the new country, both in terms of their material utility, and in terms of their total (altruistic) utility.

Proposition 7. *Consider game* G *played by two fully altruistic siblings. In equilibrium, their expected material utility is strictly decreasing in* ρ, ceteris paribus.

Thus, under our hypothesis $y_A^H = y_B^H$, the sibling pair would want to migrate if, and only if, the income riskiness in country B, $\hat{\rho}_B$, is lower than that in country A, ρ_A.

Assuming that the cost of achieving a particular success probability is the same for everyone, $\gamma_A = \gamma_B$. Would the immigrant sibling pair make a higher or a lower effort than the locals? The only difference between the immigrants and the locals is the strength of their family ties: $\alpha_A = 1 > 0 = \alpha_B$. From the above analysis we know that individuals in altruistic families may make a higher or a lower effort than selfish individuals, depending on the environment. Thus, the immigrants may get a higher or lower expected income than the locals. In any event, the immigrants enjoy a higher expected material utility than the locals.

Until now we have assumed that insurance is compulsory in country B. Suppose instead that it is voluntary, and that the immigrant sibling pair does not purchase any insurance policy. If $y_A^H = y_B^H$, Proposition 7 implies that the sibling pair would have an incentive to migrate if, and only if, $\rho_B \leq \rho_A$. Should this condition hold and the sibling pair migrate, would they necessarily be worse off than the locals, in terms of their expected material utility, by relying solely on each other's altruism rather than on the formal insurance available in country B? Surprisingly, the answer is no.

A limitation of intra-family transfers is the absence of transfers of resources from states of nature in which all family members are well off to states in which they all are poor. By contrast, an insurance market that pools the resources of a large number of individuals exposed to idiosyncratic risks enables transfers to a poor individual even when his or her sibling is also poor. However, market insurance has a disincentive effect on effort, and typically insurance companies cannot monitor effort while in many cases family members can. The question that we address here is whether this disincentive effect can be stronger than that caused by true altruism. We illustrate the relevant trade-offs between insurance and incentives in the model outlined above by comparing the expected *material* utility under optimal formal insurance with that under intra-family transfers in a few numerical examples.[9]

[9] In fact, while we put family ties at a disadvantage by assuming the smallest possible number of family members, we put it to an advantage by assuming uncorrelated risks across family members. Families that live close together and undertake identical production tasks will face positively correlated risks and hence more rarely end up with asymmetric outcomes.

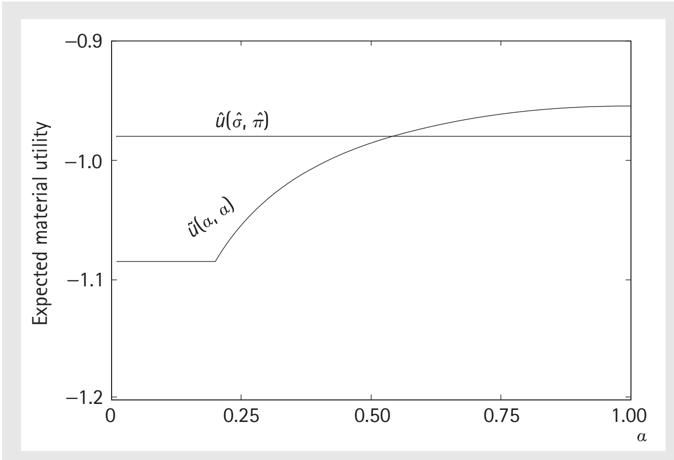


Fig. 10.4. Expected material utility under optimal insurance $\hat{u}(\hat{\sigma}, \hat{\pi})$ (horizontal line) and under true family altruism $\tilde{u}(\alpha, \alpha)$ (curve)

Consider first the case of $\rho_B = 5$ and $\gamma_A = \gamma_B = 0.5$. From section II.2, we obtain (by way of Matlab) the optimal insurance coverage $\hat{\sigma} \simeq 0.126$. This implies an equilibrium success probability of $\hat{p} \simeq 0.489$ and an expected material utility of about -0.98.[10] The horizontal line in Figure 10.4 represents this material utility level, while the curve is the graph of the expected material utility to an altruistic sibling without insurance, as a function of the common degree of altruism α. This example shows that in a society with well-functioning insurance markets, individuals with strong altruistic family ties who do not buy insurance, such as our immigrant siblings, may obtain higher expected material utility than selfish individuals who buy formal insurance, such as the natives of country B.

However, if the riskiness of the environment is sufficiently high, then formal insurance yields a higher expected material utility than fully altruistic family ties, for all degrees of altruism. With ρ on the horizontal axis and γ on the vertical, Figure 10.5 (generated with Matlab) illustrates how our fully altruistic immigrant siblings fare in comparison with the optimally insured selfish natives. In the parameter region below the curve, the siblings fare better, in terms of expected material utility, than the natives.[11]

[10] As is well known, in this type of model non-convexities in preferences over insurance policies may arise when the effect of the insurance on the effort is taken into account (see Arnott and Stiglitz 1988). However, in our examples the maximization problem is well behaved.

[11] Note that the condition $\gamma < \ln \rho$, presumed in our analysis, is met by all points in the figure.

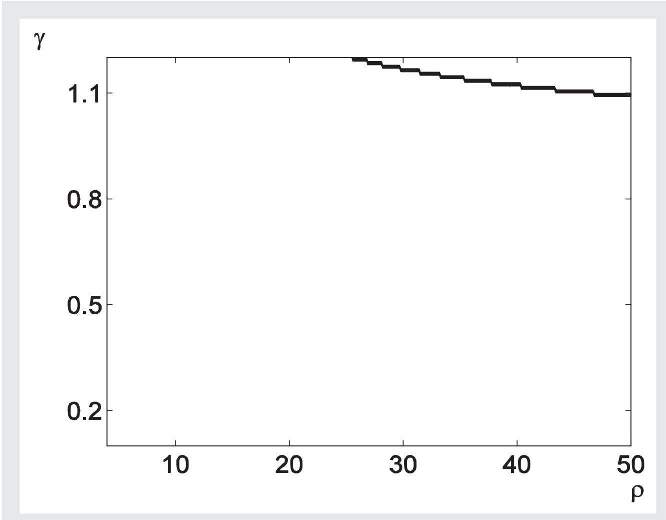


Fig. 10.5. Optimal informal insurance, $\tilde{\alpha} = \alpha = 1$, outperforms optimal formal insurance $(\hat{\sigma}, \hat{\pi})$ for values of (ρ, γ) below the curve

VI. Conclusion

Evidence indicates that family ties are stronger in some parts of the world than in others, and that this may have been so for several centuries. This observation prompted us to ask how the strength of family ties affects economic outcomes. Focusing on the family's potentially important role as an insurance provider for its members, we conducted a theoretical investigation of several questions, including how the strength of family ties affects economic outcomes such as the choice of a risk-reducing effort, and whether insurance provided within the family may be a good substitute for market insurance. To address these questions we proposed a simple model of individuals in an environment with uncertain outputs, where individuals may reduce the risk at some cost, and where families consist of pairs of identical individuals (which we may think of as grown-up siblings). In our model, insurance within a family took the form of transfers between the two family members, and although we allowed for siblings to be altruistic towards each other, we assumed that the transfers between them were dictated by social norms (coerced altruism). A parameter (ρ) captured the riskiness of the environment (the ratio between the highest and the lowest possible output).

We found that whereas coerced altruism entails a free-rider effect and thus decreases the level of risk-reducing effort, true altruism mitigates this free-rider effect by way of an empathy effect. An altruistic individual has an incentive to increase his or her risk-reducing effort for two reasons: first, to increase the probability of being

able to help a poor sibling, and second, to decrease the probability of being unlucky and forcing the sibling to help him or her out. Furthermore we showed that even if individuals are selfish, so that only the free-rider effect is present, coerced altruism is beneficial from an *ex ante* perspective, although it gives rise to involuntary transfers *ex post.*

Finally, we found that if both coerced altruism and true altruism are sufficiently strong, an individual who has access to insurance only by way of coerced altruism within the family may be better off than an individual who only has access to formal insurance. Moreover, this tends to be true if the underlying riskiness is low. If riskiness is low, selfish individuals have a low incentive to exert a risk-reducing effort even without insurance, and the moral hazard implied by formal insurance then implies that an insurance market can only play a very minor role, despite its ability to pool the resources of a very large number of individuals. By contrast, strong coerced altruism coupled with strong true altruism means that individuals have an additional incentive to exert a risk-reducing effort, namely, the utility that an altruistic individual derives from helping out the other. Where riskiness is low this empathy effect of altruism implies that individuals undertake a higher risk-reducing effort than if they were living in autarky. Thus, if individuals are rewarded for sharing with others (in our model this reward comes in the form of altruism, but society may also provide rewards), a high level of resource-sharing within the family does not necessarily imply moral hazard: indeed, *moral protection* may arise instead, whereby stronger family ties may lead individuals to choose a higher effort. Compared to market insurance, the benefit of moral protection may outweigh the cost associated with the family's inability to share the resources of a very large number of individuals.

To keep our analysis simple, we modeled a family as a pair of *ex ante* identical individuals. It would be desirable and interesting to see whether our results extend to a richer model, with larger families and heterogeneous individuals. It may also be fruitful to apply some of the concepts of our model to other settings, in particular to credit markets. In many developing countries, as well as in developed ones, we are seeing microfinance systems emerge, such as the Grameen Bank in Bangladesh (for a survey, see Armendáriz and Morduch 2005). In some of these programs poor individuals take bank loans backed by their relatives and neighbors: if a loan-taker defaults, a whole group of individuals are liable. Allowing for altruistically motivated individuals may provide additional insights into some of the successes of microfinance programs.

References

Alesina, A., and Giuliano, P. (2007), *The Power of the Family*, NBER Working Paper 13051 (Cambridge, Mass.: National Bureau of Economic Research).

Alger, I., and Weibull, J. W. (2007a), *Family Ties, Incentives and Development: A Model of Coerced Altruism*, SSE/EFI Working Paper Series in Economics and Finance 681 (Stockholm: Stockholm School of Economics).
—— —— (2007b), *The Fetters of the Sib: Weber Meets Darwin*, SSE/EFI Working Paper Series in Economics and Finance 682 (Stockholm: Stockholm School of Economics).
Armendáriz, B., and Morduch, J. (2005), *The Economics of Microfinance* (Cambridge, Mass.: MIT Press).
Arnott, R., and Stiglitz, J. E. (1988), "The Basic Analytics of Moral Hazard", *Scandinavian Journal of Economics*, 90: 383–413.
—— —— (1990), *The Welfare Economics of Moral Hazard*, NBER Working Paper 3316 (Cambridge, Mass.: National Bureau of Economic Research).
—— —— (1991), "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out of Peer Monitoring?", *American Economic Review*, 81: 179–90.
Azam, J. P., and Gubert, F. (2005), "Those in Kayes: The Impact of Remittances on Their Recipients in Africa", *Revue Economique*, 56: 1331–58.
Banfield, E. (1958), *The Moral Basis of a Backward Society* (Chicago: Free Press).
Becker, G. S. (1974), "A Theory of Social Interactions", *Journal of Political Economy*, 82: 1063–93.
Bennardo, A., and Chiappori, P.-A. (2003), "Bertrand and Walras Equilibria under Moral Hazard", *Journal of Political Economy*, 111: 785–817.
Bramoullé, Y., and Kranton, R. (2007), "Risk-Sharing Networks", *Journal of Economic Behavior and Organization*, 64: 275–94.
Broberg T., Ellingsen, T., and Johannesson, M. (2007), "Is Generosity Involuntary?", *Economic Letters*, 94: 32–7.
Bruce, N., and Waldman, M. (1990), "The Rotten-Kid Theorem Meets the Samaritan's Dilemma", *Quarterly Journal of Economics*, 105: 155–65.
Buchanan, J. M. (1975), "The Samaritan's Dilemma", in E. Phelps (ed.), *Altruism, Morality, and Economic Theory* (New York: Russell Sage Foundation).
Chami, R. (1998), "Private Income Transfers and Market Incentives", *Economica*, 65: 557–80.
Chen, Z., and Woolley, F. (2001), "A Cournot–Nash Model of Family Decision Making", *Economic Journal*, 111: 722–48.
Coate, S. (1995), "Altruism, the Samaritan's Dilemma, and Government Transfer Policy", *American Economic Review*, 85: 46–57.
—— and Ravallion, M. (1993), "Reciprocity without Commitment: Characterization and Performance of Informal Insurance Arrangements", *Journal of Development Economics*, 40: 1–24.
Cox, D., Galasso, E., and Jimenez, E. (2006), *Private Transfers in a Cross Section of Developing Countries*, CRR Working Paper 2006-2 (Chestnut Hill, Mass.: Centre for Retirement Research, Boston College).
—— and Jimenez, E. (1990), "Achieving Social Objectives Through Private Transfers: A Review", *World Bank Research Observer*, 5: 205–18.
Dana J., Cain, D. M., and Dawes, R. M. (2006), "What You Don't Know Won't Hurt Me: Costly (But Quiet) Exit in Dictator Games", *Organizational Behavior and Human Decision Processes*, 100: 193–201.
Dostie, B., and Vencatachellum, D. (2006), "Compulsory and Voluntary Remittances: Evidence from Child Domestic Workers in Tunisia", in F. Mourji, B. Decaluwé and P. Plane (eds), *Le développement face à la pauvreté* (Paris: Economica).

Edgeworth, F. Y. (1881), *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London: Kegan Paul).

Etounga-Manguelle, D. (2000), "Does Africa Need a Cultural Adjustment Program?", in L. E. Harrison and S. P. Huntington (eds), *Culture Matters: How Values Shape Human Progress* (New York: Basic Books).

Foster, A., and Rosenzweig, M. R. (2001), "Imperfect Commitment, Altruism and the Family: Evidence from Transfer Behaviour in Low-Income Rural Areas", *Review of Economics and Statistics*, 83: 389–407.

Genicot, G., and Ray, D. (2003), "Group Formation in Risk-Sharing Arrangements", *Review of Economic Studies*, 70: 87–113.

Helpman, E., and Laffont, J.-J. (1975), "On Moral Hazard in General Equilibrium Theory", *Journal of Economic Theory*, 10: 8–23.

Kimball, M. S. (1988), "Farmers' Cooperatives as Behavior Toward Risk", *American Economic Review*, 78: 224–32.

Kohler, H.-P., and Hammel, E. A. (2001), "On the Role of Families and Kinship Networks in Pre-industrial Agricultural Societies: An Analysis of the 1698 Slovenian Census", *Journal of Population Economics*, 14: 21–49.

Laferrère, A., and Wolff, F.-C. (2006), "Microeconomic Models of Family Transfers", in S.-C. Kolm and J. M. Ythier (eds), *Handbook of Reciprocity, Giving and Altruism* (Amsterdam: North-Holland).

Ligon, E., Thomas, J. P., and Worrall, T. (2002), "Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies", *Review of Economic Studies*, 69: 209–44.

Lindbeck, A., and Nyberg, S. (2006), "Raising Children to Work Hard: Altruism, Work Norms and Social Insurance", *Quarterly Journal of Economics*, 121: 1473–503.

—— and Weibull, J. W. (1988), "Altruism and Time Consistency: The Economics of Fait Accompli", *Journal of Political Economy*, 96: 1165–82.

Mendelek Theimann, N. (2006), "The Lions Mark Their Territory: The African Thought System", in K. April and M. Shockley (eds), *Diversity in Africa* (Basingstoke: Palgrave Macmillan).

Miller, D., and Paulson, A. L. (1999), "Informal Insurance and Moral Hazard: Gambling and Remittances in Thailand", unpubl. MS, Princeton University and Northwestern University.

OECD (2005), *Insurance Statistics Yearbook 1994–2003* (Paris: OECD).

Pauly, M. (1974), "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection", *Quarterly Journal of Economics*, 88: 44–62.

Posner, R. (1980), "A Theory of Primitive Society, with Special Reference to Law", *Journal of Law and Economics*, 23: 1–53.

Sen, A. K. (1977), "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, 6: 317–44.

—— (1983), "Economics and the Family", *Asian Development Review*, 1: 14–26.

—— (1987), *On Ethics and Economics* (Oxford: Blackwell).

Smith, A. (1790), *The Theory of Moral Sentiments* (London: A. Millar).

Townsend, R. M. (1994), "Risk and Insurance in Village India", *Econometrica*, 62: 539–91.

Weber, M. (1951), *The Religion of China* (New York: Macmillan).

CHAPTER 11

FROM "HARMONY" TO "COOPERATIVE CONFLICTS"

AMARTYA SEN'S CONTRIBUTION TO HOUSEHOLD THEORY

LOURDES BENERÍA

I. Introduction

I write this chapter as someone who has taught courses on Women and Work and on Gender and Development for many years, and whose research has been close to these topics. My objective is to acknowledge the ways in which Amartya Sen's work has been extremely helpful in the discussions around household analysis in our courses, and in presenting a more realistic view of how families and households function. This chapter is also written as an illustration of the ways in which Sen's work has had an enormous impact on feminist economics—from theory to empirical research and practical implications. He is one of the few prominent economists who has paid serious attention to gender inequality—particularly but not

I would like to thank Elizabeth Katz for her help with sources, and Carmen Sarasúa for very useful comments on this chapter.

exclusively from a development perspective. From his analysis of "missing women" to his work on women and development in India and his elaboration of bargaining models, his contributions have been not only substantial and well received but also very influential. His popularity among feminist economists is symbolized by the publication of a special issue of the journal *Feminist Economics* (July/November 2003) focusing on his work, and by the continuous use of his writings in teaching and research having to do with women. His work has also been taken up by women economists beyond the more strictly feminist field, such as his theoretical and empirical contributions to welfare economics and human development.

This chapter concentrates on the nature of the household and of gender relations. For many of us, the publication of Sen's "Gender and Cooperative Conflicts" (1990) represented a very important step forward in our conceptualization of household dynamics. The women's movement of the late 1960s and 1970s had been instrumental in emphasizing the focus on the household as a unit of analysis in order to understand the nature of women's work and gender inequality; it generated some of the 1970s debates on domestic labor and the family that had much influence on the economic analysis of the household. To be sure, the New Household Economics had pioneered this analysis during the 1960s—with roots in Jacob Mincer's work in the 1950s—with the objective of understanding home production, home consumption, and the gender division of labor as well as the factors affecting women's (and men's) labor force participation. Beginning with Mincer's and Gary Becker's work, many other authors in the 1960s and 1970s developed neoclassical models in different directions. They were used to analyze a variety of topics ranging from explaining time use and the gender division of labor to marriage patterns and fertility rates (see e.g. Lloyd 1975; Lloyd and Niemi 1979). However, parallel to the New Household Economics, other analytical models of the family and the household sprang from non-mainstream feminist perspectives during the same period (e.g. Humphries 1977; Hartmann 1987); although their weight within the economics profession was of course much less significant than that of neoclassical models, they were important from a more interdisciplinary perspective. In what follows, I analyze the differences between these conceptual frameworks in order to provide the background that helps contrast them with Sen's conceptualization of the household as an area of "cooperative conflicts".

II. The Harmonious Household: Neoclassical Models and their Critics

Much has been written about the extent to which neoclassical models have been built around the notion of the unitary household making decisions based on a

single utility function and without differentiating between different household members. In its typical formulation, utility maximization takes place with regard to consumption, production and general time use, having an impact on a wide range of decisions such as those affecting participation in the paid labor market by different household members, their leisure time, fertility rates, and so on. Despite the acceptance and wide use of these models in the profession as powerful tools of analysis, critics have noted a variety of weaknesses: (1) the questionable assumption that family well-being can be efficiently maximized by a "benevolent dictator" who decides altruistically on resource allocation and oversees the processes of "specialization" around the division of labor—in and outside the household; (2) the inability of such conceptualizations to capture differences and tensions among household members around these decisions; (3) the failure of utility maximization analysis in terms of questioning the initial distribution of household members' skills, which are taken as given; (4) the neglect of the negative consequences of specialization in the traditional domestic division of labor for women; (5) the assumption of "given preferences" which are affected by prevailing social norms but also by individual agency and contestation of these norms and by social change; (6) the implication of inevitability of outcomes by neoclassical equilibrium analysis. These and other critiques are well known and have been provided in detail by many authors, particularly feminist economists writing from a concern with gender inequality and women's oppression (Folbre 1988, 2001; Katz 1991; Koopman 1991; England 1993; Benería 1995; Bergman 1995; Nelson 1995). Sen himself (1982, 1983, 1985) has criticized the use of household models based on economic rationality. In what follows, I will focus mostly on the notion of the "harmonious household" as the central concept in neoclassical models characterizing the family and household economic decision-making.

In order to model a functioning, efficient and maximizing household, neoclassical analyses needed to solve the problem of assuming a unitary utility function encompassing various individuals whose tastes and objectives might differ substantially. This was attempted either with the explicit assumption of harmony around household decision-making or, in the case of "rotten kids" who might dissent, with the assumption of an altruistic household head who can arrive at the most efficient way to maximize household welfare (Becker 1981). Harmony then implies either total agreement on decisions affecting household members of different genders and ages, or an authoritarian though altruistic household head who makes decisions despite what other members, even "rotten kids", might prefer. Given what we know about families and how they function, the high level of harmony required in the first case is a rather unrealistic assumption; in the second case, the benevolent "dictator" reflects the functioning of the traditional patriarchal family, in which gender relations are bound by an authoritarian and patriarchal structure. This is quite different from the achievement of "harmony" and leaves us with the understanding that neoclassical models do not question the unequal gender relations that characterize

patriarchal households. In addition, these models can justify, in the name of economic rationality, highly discriminatory behavior against girls and women. Examples of such discrimination abound across cultures, such as dowry systems that result in female infanticide because girls are viewed as generating extra costs associated with the dowry; or in cases when girls' educational levels are deliberately kept lower than boys', based on assumptions and social norms concerning male and female responsibilities and what constitutes the "proper" gender division of labor.

Thus, the appearance of "harmony" can result from the acceptance of social norms and regulations that function as "exogenous dictators" and act upon the given "tastes" and "preferences" assumed in neoclassical models. In this sense, maximization might rest on institutions and traditions that are oppressive to women and to other household members, such as young males subject to their father's authority. This acceptance on the part of those discriminated against might be *real* and reflect a form of false consciousness, or it may only be *apparent* and result from fear of the consequences of non-compliance; or it might reflect a lack of agency. In all these cases, the appearance of harmony in household decision-making hides the asymmetries of unequal gender (and age) relations. Building theoretical models on these bases contributes to their legitimization and perpetuation. It is no wonder that neoclassical models have been criticized by those who object to their implications for gender inequality. To be sure, some economists using these models have become sensitive to the various critiques and have taken them, as well as alternative views, into consideration (Blau, Ferber and Winkler 2006). Yet, a good part of the profession has ignored these critiques and has continued to rely on the assumption of a unitary household.[1] Becker himself paid no attention to this issue in his influential *A Treatise on the Family*, even though the critiques were well known when the book appeared in 1981.

As for the neglect of the importance of social institutions and norms that influence household dynamics, this excludes important factors that can vary significantly across countries. Empirical research has shown that neoclassical models misrepresent the ways in which many households function, particularly in non-Western cultures. For example, Wolf's work in the 1980s on Java and Taiwan showed how decisions related to the labor allocation of family members can vary across cultures. Her empirical work showed that, in Taiwan, parents exercised a strong influence on their daughters' participation in wage work, whereas in Java, daughters contested their parents' decisions to a much greater extent (Wolf 1990). Similarly, Koopman's extensive study of agricultural households in southern Cameroon and elsewhere in Africa led her to argue that, rather than operating as a single unit of production, in most cases the male household head and his wife or wives conducted "separate enterprises", earned separate incomes, and managed separate budgets. To be sure, men and women did often work together under a "manager who

[1] For a recent textbook example, see Bryant and Zick (2006).

mobilizes inputs and controls output". However, this "managing" took different forms and involved both men and women (Koopman 1991). A different example was provided by Katz's study on rural Guatemala, in which she examined the "complex processes" that characterize the domestic economy. More specifically, she showed the "separateness of the male and female economic spheres and their linkages via transfers of income, land, goods and labor" (Katz 1995: 329). She concluded that this complexity implies both individual autonomy and mutual exchange. Likewise, Bina Agarwal's work has examined extensively the importance of social norms and cultural practices in shaping the ways in which households regularly operate (see below). Finally, the anthropological literature has added further evidence to the notion that, in many cultures, the unitary household is not the best model of the ways in which they make choices and organize themselves on a daily basis (Guyer 1983; Geisler and Hansen 1994). More importantly, recent conceptualizations of the household extend this notion of a non-unitary household to a more general model applicable to different cultures, including Western countries (Carter and Katz 1992; Jacobsen 2007).

A different but related critique of neoclassical theory results from the contradiction between what England (1993) called the "separative self model" of individual economic rationality and the altruism assumed within the household. As England argued, choices and decisions based on economic rationality assume autonomous individuals without emotional connections to each other such as those resulting from feelings of love, empathy, anger or antagonism. Yet, household theory *à la* Becker relies on an altruistic patriarch able to make decisions based on a single family utility function—hence the contrast between choices and behavior within the domestic realm and in the labor market. England called the assumptions behind neoclassical household theory "androcentric" or male-centered, biased in favor of men's interests and based on a male-dominated system of gender relations, as I argued above.

III. Left-Feminist Approaches: The Working-Class Family, the Conflictive Household, and Exploitation in the Home

During the 1970s, left-leaning analysis dealt with quite different concepts of the household, much less known among orthodox economists, grounded in Marxian and feminist theory. A variety of authors contributed to the conceptualization

of the family from this non-orthodox perspective (Delphy 1976; Humphries 1977; Himmelweit and Mohun 1977; Hartmann 1979, 1987; Molyneux 1979; Folbre 1982). Their views illustrate the tensions between feminism and the more traditional Marxian analysis, which ignores gender differences and inequality within the household. In comparison with neoclassical models, these approaches tended to ask wider and more political questions regarding the nature of the family and the gender division of labor, particularly in terms of the function of these within the capitalist economy. They also pointed to the household and to women's domestic and reproductive work as a central focus of women's oppression.

An initial effort to understand the economic significance of women's work at the domestic level and also in the larger economy appeared in the late 1960s and 1970s as a result of the questions raised by the second wave of feminism. It is beyond the scope of this chapter to provide a comprehensive list of the different contributions that generated the "domestic labor debate" in the 1970s. Instead, in what follows, I focus on two contrasting positions, typified by Humphries (1977) and Hartmann (1979, 1987), which provide very different notions of the family and are useful to compare with Sen's work. Subsequently, Folbre's analysis (1982), also using a Marxian framework, raised interesting questions with regard to the notion of "gender exploitation"—as opposed to class or capitalist exploitation—and bargaining power among different household members. As in the case of neoclassical models, a retrospective look at this early literature helps us to understand the subsequent contributions of bargaining models and of Sen's notion of cooperative conflicts.

Humphries' 1977 article was a response to a functionalist Marxian view of the family that viewed it as the arena of domestic labor serving the needs of capitalism, particularly through the maintenance and reproduction of the labor force. Her 1987 article was an effort to "redress this bias" by arguing that, in certain periods of capitalist development, working-class families presented a common front—quite apart from the needs of capital and even in conflict with it—as a result of their struggle for survival. Her historical analysis from nineteenth-century England showed how working-class families functioned as a unit in order to preserve family ties that contributed to their ability to face problems of survival. Humphries' objective was to illustrate how "family ties, vitalized by ideology, bind together labouring and non-labouring individuals" (28) because "in nineteenth century British capitalism kinship ties provided a major source of non-bureaucratic support in conditions of chronic uncertainty" (31). Hence she suggested that, at least in nineteenth-century Britain, the endurance of the family and traditional kinship ties among working-class families persisted, and family loyalties sustained many individuals through the turbulent period of industrialization.

In terms of gender relations, Humphries discussed women's position within the family and as wage laborers. However, she focused on unequal gender relations only when discussing men's and trade unions' defense of the family wage—which

resulted in a decrease in women's employment. She added that the campaign for a family wage was not entirely disadvantageous for working women, first because, as also argued by other authors, men were the higher earners in the household while women were at the lower end of the labor hierarchy, and thus their wages were lower. Second, women also benefited from the family wage as members of a household. Although Humphries mentioned the existence of "patterns of dominance relating to age, sex and division of labour", she argued that they "should not blind observers to the material benefits that the family imparts to the working class in its struggle for a better life" (39). Likewise, although she later explored the ways in which male privilege in the labor market was historically constructed in specific cases (Humphries 1980), she did not focus on gender-related hardships and tensions that family loyalties might have imposed upon some household members, and upon women in particular. To be sure, she did not necessarily assume a "harmonious household" *à la* neoclassical theory, but her view of the family emphasized class solidarity rather than gender inequality. In this sense, it differed significantly from feminist models focusing on "disharmony" around gender-related tensions (see below). In fact, her argument could be read as a critique of such models, since it was an attempt to explain why "working class women might reject a [women's] movement that, in effect, denies them any semblance of self-determination and dignity" (Humphries 1977: 36).

A different issue is the extent to which it is possible to generalize from Humphries' work, since her empirical analysis relates only to working-class families and to nineteenth-century developments in England. Along these lines, her arguments can be applied to other social settings and periods. An example is provided by the literature that emerged from studies on the impact of structural adjustment policies on household survival strategies during the late 1980s and 1990s. Studies focusing on the social costs of adjustment in developing countries showed the ways in which poor households in particular had to pool their resources in order to survive the belt-tightening policies that generated budget cuts, unemployment, the deterioration of labor conditions, and lack of social protection for a large proportion of the population.[2] Some authors used notions such as "privatization of survival" to refer to the enormous efforts made by households to deal with these difficult conditions without social protection,[3] thus echoing Humphries' view of the working-class family and of kinship ties as a source of endurance through turbulent periods. However, the literature also documented the extent to which structural adjustment policies were particularly hard on women; the multiple gender dimensions were illustrated by the intensification of their work within the home and in the labor market as well as their increased need to participate in

[2] For summaries of this literature, see Benería (1999), González de la Rocha (2000) and Pérez-Sáinz (2005), among others.

[3] See Benería and Feldman (1992).

community services and other forms of survival.[4] This research drew attention not only to the existence of gender inequalities but also to the complexities of household relations.

In contrast with Humphries' analysis, Hartmann's emphasis centered on the family as a locus of conflict and struggle. While pointing out that research on families had contributed enormously to our understanding of diversity in family structures and relationships, Hartmann argued that this research "generally failed to identify and address sources of conflict within family life", thus limiting its usefulness for understanding women's situation (Hartmann 1987: 110). Historians, she argued, had generally seen the family as a source of dynamic change and as a key actor and agent of social change, often assuming a unity of interests and downplaying conflicts among family members. To be sure, Hartmann also recognized that family historians had documented tensions between households and the world outside them, thereby suggesting that, as Humphries showed, they can act as unified entities; this was partially the result of the mutual dependency among household members. On the other hand, Hartmann criticized traditional Marxian analysis for having failed to follow Engels' project of understanding the two arenas in "the production and reproduction of immediate life", namely, the production of "means of existence" and that of "human beings themselves". Their concentration on the production of "means of existence" outside the household led them to neglect the area of reproduction of human beings. Thus, Hartmann's feminist analysis focused on the conflicts and tensions that result from the nature of the household as the locus where reproduction, production and distribution take place—mediated by the gender division of labor and by the centrality of domestic work for women.

Hartmann's basic argument was that, as capitalism developed, men's concentration on the sphere of paid production and public life and women's concentration on the household contributed to gender differentiation. Although households continued to make decisions related to issues such as wage work, pooling income, having children and caring for dependent members, men tended to have more power than women. Thus, viewing the household as a unit obscured the reality of unequal gender relations; for Hartmann, the family is the primary arena where "men exercise their power over women's labor". This inequality, she pointed out, can take extreme forms since "mutual dependence by no means precludes the possibility of coercion". Similarly, she pointed out that "women and men are not less mutually dependent in the household than are workers and capitalists or slaves and slave owners" (Hartmann 1987: 117). Her analysis of time spent on housework in the US led her to argue that women of all classes performed household labor for men; despite changes in women's participation in the labor

[4] In the Andean countries, for example, poor women played a crucial role in organizing *comedores populares* or collective soup kitchens that often fed entire neighborhoods during the economic crises of the 1980s. This intensified their workload along the lines of traditional gender roles.

force, the prospects for change, she argued, did not point towards equal sharing of housework between men and women. Hence the importance of focusing on unequal gender relations within the household in order to understand gender inequality.

Along these lines, Folbre (1982) reinforced Hartmann's basic argument, with an interesting attempt to apply the Marxian concept of exploitation to the economics of the household. She used a model of household production and distribution to argue that labor exchanges among family members might be unequal to the extent that some family members, such as those concentrating on domestic work, can be "exploited" by those engaged in wage work. At the same time, she argued, access to income in the labor market and the various degrees of economic independence contribute to the differences in bargaining power among family members. Consequently, she added, men may enjoy substantially greater bargaining power than women and children. Folbre applied her theoretical insights to the analysis of demographic transitions, arguing that the cost of children and the increasing importance of education leads to fertility declines and to changes in the relative bargaining power of women and children. Similarly, an increase in women's employment might increase their bargaining power, although its immediate effect is not clear due to the persistence of other forms of gender inequality. In this sense, she emphasized the importance of institutions and patriarchal forms affecting outcomes at the household level. Her analysis, like Hartmann's, tended to emphasize the notion of the family as an institution where exploitation and gender inequality can prevail, but she also pointed to factors tending to generate change in the relative position of household members.

Similarly, and despite Hartmann's emphasis on the family as a locus of conflict based on gender, she recognized that anthropological and historical research focusing on households revealed the family "as an embodiment of both unity and disunity", a view that contrasts with the main thrust of her analysis. Thus, in the Marxist-feminist theoretical framework of the period, Hartmann's and Folbre's approaches reflected strong feminist views giving first priority to gender divisions. On the other hand, Humphries' view of the family was more influenced by the Marxist's emphasis on class divisions than by the feminist focus on relations of domination/subordination between men and women. The tension between these two approaches has far from disappeared over the years but, as illustrated by the work on the effects of structural adjustment policies mentioned above, the two tended to converge while de-emphasizing the ideological aspects of the earlier period. This is particularly so with the new conceptualizations of gender analysis developed since the mid-1980s.[5] With regard to the concept of family relations, this meant that the household can be both harmonious and conflictive. More recent feminist household theory, including bargaining models, evolved in this direction.

[5] For an elaboration of this argument, see Benería (2003: ch. 2).

IV. Cooperative Conflicts

The bargaining approach represented a very important step forward in the conceptualization of the family in economic analysis. Given the above critiques of neoclassical models and the shortcomings of early feminist approaches, bargaining models were able to incorporate precisely the notion that households can be both harmonious and conflictive, thus introducing a very useful way of understanding family dynamics (Sen 1983, 1990; Katz 1991, 1997; Lundberg and Polak 1993; Kanbur and Haddad 1994; Agarwal 1994, 1997). After rejecting the predominant neoclassical theory and finding other models wanting, the challenge was the construction of an alternative way to explain how households make decisions and organize themselves. This is particularly true in terms of the *processes* involving time use, the gender division of labor, and allocation of income and other resources among household members. As Katz very appropriately pointed out, a feminist household model would "put process in the forefront, while not abandoning the theoretical task of explaining the observed outcomes of household resource allocation" (Katz 1991: 45).

In particular, Sen's notion of the family as a locus of "cooperative conflicts" captures the essence of the contradictions and tensions within households, and was a significant step forward from previous conceptualizations. His 1990 article was very instrumental in the effort to integrate gender into bargaining models. His point of departure was the reluctance on the part of most economists in the standard literature on economic development to include gender. Responding to the calls that many women had voiced since Ester Boserup wrote her pioneer book on *Woman's Role in Economic Development* (1970), he emphasized the systematic "inferior position of women inside and outside the household in many societies ... [which points] to the necessity of treating gender as a force of its own in development analysis" (Sen 1990: 123). The fact that the relative deprivation of women *vis-à-vis* men is not uniform across countries, he argued, "was an important reason for giving serious attention to the causal antecedents of the contrasting deprivations" (124). The issue of "missing women", particularly in Asian countries, was another point of departure for considering the importance of gender in development. However, Sen recognized the difficulty of translating this awareness into "an adequate framework for the use of gender categories and sex-specific information in social analysis" and he found the problem "too complex to be resolved by any kind of simple model" (124). His approach to the problem consisted in integrating development and gender issues into the framework of bargaining models on the basis that they could capture "the coexistence of extensive conflicts *and* cooperation in household arrangements" (125). For those of us who found the assumptions of either the harmonious or the conflictive household laking complexity, incomplete or politically questionable, the notion of cooperative conflicts seemed to open up

a very promising avenue for household theory and gender analysis, for a variety of reasons.

First, Sen's model began with his analysis of capabilities, well-being, agency and perceptions but also with his understanding of the family and identity as exercising "such a strong influence on our perceptions that we may not find it easy to formulate any clear notion of our own individual welfare" (125). This is of course a theme that Sen had developed in earlier writings (see Sen 1982, 1984) and it is particularly relevant in thinking through bargaining processes. The problem of inadequate perceptions and false consciousness is especially relevant for women, as a result of gender constructions leading to a weak understanding of personal welfare. Sen saw clearly how this understanding is shaped by cultural norms, perceptions of self and of others, ability to articulate personal needs, and even fear of ostracism and of other negative consequences—such as male violence—if prescribed and internalized norms are contested or broken. In fact, Sen saw these problems as applying in a more general way to deprived groups that might be "habituated to inequality", "unaware of possibilities of social change", "hopeless" regarding social change, "resigned to fate", and even "willing to accept the legitimacy of the established order". Overcoming these problems is the essence of human development and points to the importance of Sen's capability theory. In addition, his analysis opened the door to an inclusion of cultural factors and of differences across countries in our understanding of gender relations.

Second, by focusing on a "social concept of technology", Sen emphasized the importance of social organization that "permits the use of specific techniques of production" and applied it to the household, where he saw both cooperation and conflict at work. By doing so, Sen paid attention to the importance of *processes* in household dynamics, including the process of negotiation over the division of labor and resource allocation. He saw social technology as reflecting social arrangements regarding production in a broadly defined way, for example incorporating cultural aspects, and not merely maximizing objectives, in the analysis of the gender division of labor and the factors affecting it. This allows the "deeper probing" that is instrumental in clearing away "the fog and ambiguity in which the roles of different types of laboring activities are hidden by stereotyped social perceptions" (129). Similarly, "social technology" shapes gender-related distribution of resources affecting both the economic survival of families and social/cultural distributional parameters.

Third, by building on earlier models of cooperative and non-cooperative bargaining, Sen re-examined the tendency of orthodox economics to ignore the simultaneous existence of cooperation and conflict in households; the more realistic assumption of interaction between the two leads to a focus on the bargaining process. However, his awareness of the ambiguities both of perception of interests and of notions of legitimacy led him not only to examine the bargaining problem itself but also that of the constant interaction and links between perceptions,

responses, well-being and agency. Sen's 1990 article made clear that the relative bargaining power of individuals depends on their *breakdown position* but also on their *perceived interest* and *perceived contribution* responses. These concepts are rich with implications for gender analysis; we were of course aware that women tended to be weak bargainers but Sen's work took us much closer to understanding why, and to capture the contributing factors with more detail. However, as in the case of other authors who had used game theory to build household models, he did not clearly focus on gender-related power relations even though he laid the basis for further elaboration. It is for this reason that Sen's work has often not been seen as "feminist"—in the sense of addressing unequal gender relations. However, this is in fact a view that Sen has contested, pointing out the extent to which "power" is directly involved in his various assessments of gender inequality—from his analysis of capabilities to that of cooperative conflicts. Regarding the latter, his response to this question has been to emphasize that

> [i]n the emergence of some cooperative solution *among the many that are available*, the powers of the two parties play a crucial part; for example, the more powerful party can obtain more favorable divisions of the family's overall benefits and chores.
>
> (Sen 2003: 324; emphasis mine)

Indeed, Sen's view of the cooperative solution "among the many that are available", represented a most interesting contrast with the neoclassical "equilibrium" solution, which conveys the notion of inevitability of outcome. Indeed, the importance of power relations for the cooperative outcome in bargaining is implicit in the model. What Sen did not incorporate was the detail in the nature of gender relations, which has been further elaborated by feminist economists (see below).

Fourth, by pointing out that "conflicts of interest between men and women are unlike other conflicts, such as class conflicts" (Sen 1990: 124), Sen reinforced arguments concerning the importance of the "personal" in feminist discourses. He argued that, unlike men and women, "a worker and a capitalist do not typically live together under the same roof... Thus, this aspect of 'togetherness' gives the gender conflict some very special characteristics" (Sen 1990: 147). Tensions and conflicts within the household may develop "against the background of pervasive cooperative behavior" (Sen 1990: 127). This opened the door to the inclusion of material interests as well as non-material aspects of gender relations in the bargaining process. Sen emphasized the importance of paying attention to "*perception problems* about respective interests, contributions and claims" (Sen 1990: 130; emphasis mine). Hence the importance placed on the "breakdown well-being response," the "perceived interest response" and the "perceived contribution response" in his analysis (Sen 1990: 126–30). Further, this opened the door to understanding the outcome of bargaining processes by considering factors related to gender constructions that form part of individual characteristics—such as self-esteem and feelings of subordination/domination. An emphasis on the role of perceptions also opens the

door to overcoming the problems derived from England's "separative" household model, since perceptions contribute to understanding the emotional and other links between household members.

Fifth, Sen's analysis was very conducive to thinking through policy and action that would improve women's bargaining power and consequently their condition. I still remember the first time I assigned his 1990 article in one of my courses; the discussion led very easily to the topic of concrete practical actions and policies in a way that other theoretical models did not. Students saw clearly the connections between, for example, the low self-esteem typical of many women and the possibility of improving it through education, access to income, or other empowering mechanisms. This opened up further discussion on the different possibilities through which educational programs and participation in the paid labor force could function as a way of improving women's "fallback position" or their "perceived interest" responses. Similarly, it was easy to formulate the role that women's organizations and increased agency could play in making development relevant for women. Sen brought up many of these issues in his work on India's economic development (Drèze and Sen 1995), an interesting and very useful book not only for its empirical content but also for its importance in legitimizing gender analysis in development. Further, and beyond his 1987 article, Sen's capabilities approach reinforced the usefulness of his analysis for thinking through gender-aware social policy.

It was no surprise that feminist economists welcomed Sen's work but they also extended it in different ways, insisting particularly on the importance of the wider set of factors affecting gender inequality beyond the narrowly defined economic variables. To illustrate, Elizabeth Katz (1997) built on Sen's work and, more generally, enriched bargaining models in different ways. First, she pointed out that one of the shortcomings of these models is their symmetric treatment of family members as players with regard to "voice"—the right and ability to participate in the household bargaining process. Similarly, this symmetric treatment also applies to "exit"—the ability to perceive and fall back on available alternatives to a cooperative solution. Far from being gender-symmetric, she argued, voice and exit can be very different for men and women. The difference can be related to a variety of factors, ranging from valuation of one's and others' earnings to social norms and cultural practices, to the existence of alternatives that can replace the prevailing gender-differentiated social sanctions for a partner breaking marriage bonds. Yet, it is worth noting that, although Katz's contribution clearly moved this type of analysis to a new level, Sen had underlined the importance of some of these factors in determining fallback positions and thread points, and he had provided examples in his 1990 article.

Another illustration of how Sen's concepts have been elaborated and extended by feminist economists is provided by the work of Bina Agarwal (1994, 1997) on the role of social norms and cultural practices as determinants of intra-household

bargaining processes. What determines household members' bargaining power, she argued, had tended to be too narrowly defined as a result of concentrating on economic factors and not exploring qualitative aspects of power and power relations. She made use of her rich empirical information, mostly from India and other South Asian regions, to analyze different arenas in which household bargaining takes place. Among the relevant factors examined were the subsistence needs within the family, social perceptions of a person's contribution to family resources, effective command over land property and land rights, and social norms that had mostly been treated as exogenous (except by mostly feminist economists). Agarwal's work filled the many gaps regarding the specificity of social norms and their functions in different cultural contexts. They can, she argued, (1) set the limits of what can be bargained; (2) determine the constraints on bargaining power; (3) affect how bargaining is conducted; and (4) generate bargaining over social norms themselves when norms are challenged. The detail provided by interdisciplinary empirical information made her analysis particularly rich and thorough. Agarwal also extended her investigation beyond the household and examined the extent to which extra-household parameters impinge upon intra-household bargaining and vice versa. Such is the case with the labor market, the community and the state, all of which she analyzed separately and in their interaction with household bargaining. As she points out, many of the factors affecting household bargaining could be incorporated into formal models, quantified and empirically tested, while some others refer to "qualitative dimensions on which systematic information is often difficult to gather, and/or which cannot be easily integrated into formal models" (Agarwal 1997: 37), as in the case of perceptions.

In sum, Sen's notion of "cooperative conflicts" was instrumental in throwing light on much that had been left unexplored or under-analyzed in the "black box" of the household. As a result, we have learned a great deal about the factors contributing to household bargaining. Prior to this illumination (and even afterward in some cases), feminist analysis of household dynamics had implicitly referred to some form of bargaining process but without an explicit formulation of a model (Pastner 1974; Benería and Roldán 1987; Mencher 1988; González de la Rocha 2000). As argued above, one of Sen's important contributions is the view of bargaining as a *process* driven by implicit or explicit *strategies* leading to a multiplicity of outcomes. Similarly, he opened the door to the inclusion of a wide range of non-economic factors affecting bargaining and the relative power of household members. In doing so, he has been instrumental in clarifying our notions about the nature of the family and in improving our understanding of gender dimensions in household theory. Finally, his analysis has been very conducive to formulating action and policy. To be sure, some aspects of Sen's work were built on existing analysis contributed both by neoclassical and unorthodox economics, including feminist economists, but he moved their work forward and his authority gave it more legitimacy and visibility. Its impact has not diminished with time.

References

Agarwal, B. (1994), *A Field of One's Own: Gender and Land Rights in South Asia* (Cambridge: Cambridge University Press).

—— (1997), "Bargaining and Gender Relations Within and Beyond the Household", *Feminist Economics*, 3: 1–51.

Becker, Gary S. (1981), *A Treatise on the Family* (Cambridge: Cambridge University Press).

Benería, L. (1996), "Toward a Greater Integration of Gender and Economics", *World Development*, 23: 1839–50.

—— (1999), "Structural Adjustment Policies", entry in J. Peterson and M. Lewis (eds), *The Elgar Companion to Feminist Economics* (Cheltenham and Northampton, Mass.: Edward Elgar), 687–95.

—— (2003), *Gender, Development and Globalization: Economics as if All People Mattered* (New York: Routledge).

—— and Feldman, S. (1992), *Unequal Burden: Economic Crises, Persistent Poverty and Women's Work* (Boulder, Colo.: Westview Press.

—— and Roldán, M. (1987), *The Crossroads of Class and Gender: Homework, Subcontracting and Household Dynamics in Mexico City* (Chicago: University of Chicago Press).

Bergman, B. (1995), "Becker's Theory of the Family: Preposterous Conclusions", *Feminist Economics*, 1: 141–50.

Blau, F., Ferber, M., and Winkler, A. (2006), *The Economics of Women, Men and Work*, 5th edn (Upper Saddle River, NJ: Pearson Prentice Hall).

Bryant, W. K., and Zick, C. (2006), *The Economic Organization of the Household* (Cambridge: Cambridge University Press).

Carter, M., and Katz E. (1997) "Separate Spheres and the Conjugal Contract: Understanding the Impact of Gender-Biased Development", in L. Haddad, J. Hoddinott and H. Alderman (eds), *Intrahousehold Resource Allocation in Developing Countries: Models, Methods and Policy* (Baltimore: Johns Hopkins University Press), 95–111.

Delphy, C. (1976), *The Main Enemy* (London: Women's Research and Resource Centre).

Drèze, J., and Sen, A. K. (1995), *India: Economic Development and Social Opportunity* (Delhi: Oxford University Press).

England, P. (1993), "The Separate Self: Androcentric Bias in Neoclassical Assumptions", in M. Ferber and J. Nelson (eds), *Beyond Economic Man* (Chicago: University of Chicago Press), 37–53.

Folbre, N. (1982), "Exploitation Come Home: A Critique of the Marxian Theory of Family Labor", *Cambridge Journal of Economics*, 6: 317–92.

—— (1988), "The Black Four of Hearts: Towards a New Paradigm of Household Economics", in J. Bruce and D. Dwyer (eds), *A Home Divided* (Stanford, Calif.: Stanford University Press), 248–64.

—— (2001), *The Invisible Heart: Economics and Family Values* (New York: New Press).

Geisler, G., and Hansen, K. T. (1994), "Structural Adjustment, The Rural–Urban Interface and Gender Relations in Zambia", in N. Aslanbeigui, S. Pressman, and G. Summerfield (eds), *Women in the Age of Economic Transformation* (London and New York: Routledge), 95–112.

González de la Rocha, M. (2000), "Private Adjustments: Household Responses to the Erosion of Work", UNDP/SEPED Conference Paper Series, available at <hdr.undp.org/docs/events/global_forum/2000/rocha.pdf>.

Guyer, J. (1983), *Anthropological Models of African Production: The Naturalization Problem* (Boston: African Studies Center, Boston University).

Hartmann, H. (1979), "The Unhappy Marriage of Marxism and Feminism: Towards a More Progressive Union", *Capital and Class* 8 (Summer): 1–33.

—— (1987), "The Family as the Locus of Gender, Class, and Political Struggle: The Example of Housework", in S. Harding (ed.), *Feminism and Methodology* (Bloomington, Ind. and Indianapolis: Indiana University Press), 109–34.

Himmelweit, S., and Mohun, S. (1977), "Domestic Labour and Capital", *Cambridge Journal of Economics*, 1: 15–31.

Humphries, J. (1977), "The Working Class Family, Women's Liberation, and Class Struggle: The Case of Nineteenth Century British History", *Review of Radical Political Economics*, 9: 25–53.

—— (1980), "Protective Legislation, the Capitalist State, and Working Class Men: The Case of the 1842 Mines Regulation Act", *Feminist Review*, 7: 1–33.

Jacobsen, J. (2007), *The Economics of Gender* (Oxford: Blackwell).

Kanbur, R., and Haddad, L. (1994), "Are Better Off Households More Unequal or Less Equal?", *Oxford Economic Papers*, 46: 445–58.

Katz, E. (1991), "Breaking the Myth of Harmony: Theoretical and Methodological Guidelines to the Study of Rural Third World Households", *Review of Radical Political Economics*, 23(3–4): 37–56.

—— (1995), "Gender and Trade within the Household: Observations from Rural Guatemala", *World Development*, 23(2): 327–42.

—— (1997), "The Intra-Household Economics of Voice and Exit", *Feminist Economics*, 3(3): 25–46.

Koopman, J. (1991), "Neoclassical Household Models and the Modes of Household Production: Problems in the Analysis of Agricultural Households", *Review of Radical Political Economics*, 23(3–4): 148–73.

Lloyd, C. (1975), *Sex, Discrimination and the Division of Labor* (New York: Columbia University Press).

—— and Niemi, B. (1979), *Economics of Sex Differentials* (New York: Columbia University Press).

Lundberg, S., and Polak, R. (1993), "Separate Spheres Bargaining and the Marriage Market", *Journal of Political Economy*, 101: 988–1010.

Mencher, J. (1988), "Women's Work and Poverty: Contribution to Household Maintenance in Two Regions in South India", in D. Dwyer and J. Bruce (eds), *A Home Divided: Women and Income in the Third World* (Stanford, Calif.: Stanford University Press), 99–119.

Molyneux, M. (1979), "Beyond the Domestic Labour Debate", *New Left Review*, 116: 3–27.

Nelson, J. (1995), "Economic Theory and Feminist Theory: Comments on Chapters by Polachek, Ott, and Levin", in E. Kuiper and J. Sap (eds), *Out of the Margin: Feminist Perspectives on Economics* (London and New York: Routledge), 120–5.

Pastner, C. M. (1974), "Accommodations to Purdah: The Female Perspective", *Journal of Marriage and the Family*, May: 408–14.

Pérez-Sáinz, J. P. (2005), "Labor Exclusion in Latin America: Old and New Tendencies", in N. Kudva and L. Benería (eds), *Rethinking Labor Market in Formalization: Poverty, Precarious Jobs and Social Protection* (Ithaca, NY: Internet First University Press), 67–85.

Sen, A. K. (1982), *Choice, Welfare and Measurement* (Oxford: Blackwell; Cambridge, Mass.: MIT Press).

SEN, A. K. (1983), *Resources, Values and Development* (Oxford: Blackwell; Cambridge, Mass.: Harvard University Press).
—— (1985), "Women, Technology and Sexual Divisions", *Trade and Development*, 6: 195–223 (New York: United Nations).
—— (1987), *On Ethics and Economics* (Cambridge, Mass.: Blackwell).
—— (1990), "Gender and Cooperative Conflicts", in I. Tinker (ed.), *Persistent Inequalities: Women and World Development* (New York: Oxford University Press), 123–49.
—— (2003), "Continuing the Conversation", *Feminist Economics*, 9(2): 319–32.
WOLF, D. (1990), "Daughters, Decisions, and Domination: An Empirical and Conceptual Critique of Household Strategies", *Development and Change*, 21: 43–74.

CHAPTER 12

FAMINE, WIDOWHOOD AND PAID WORK

SEEKING GENDER JUSTICE IN SOUTH ASIA

MARTHA ALTER CHEN

In the fall of 1987, I heard Amartya Sen speak for the first time. He gave a talk on human development at the Harvard Institute for International Development (HIID), which I had joined earlier that year. The small conference room was jammed with students and faculty. After Sen's talk, an HIID colleague opened the discussion with the following question: "Amartya, aren't you being a little hard on GDP as a measure of development?" Sen replied: "I am not questioning whether GDP *can* be used as a measure of development but whether it *should* be used." This was my first experience of Amartya Sen in person—his remarkable combination of intelligence, wit and compassion.

At the time, I knew Amartya Sen only by reputation. Little did I know that my husband Lincoln and I would become close friends and colleagues of Amartya-da. What I did know was that his approach to economics appealed to me. Since joining

HIID, I had been struggling for some months with many of the assumptions of my neo-classical economic colleagues. In Amartya-da, I found not just a friend but a mentor from whom I learned that there are alternative approaches to economics—more normative and justice-based—and who took great interest in what interested me: the livelihood and coping strategies of poor households—and especially of the women in those households—in South Asia.

In this chapter, I will explore gender injustice in three seemingly distinct, but related, contexts in South Asia: famine, widowhood and paid work. Amartya Sen has written extensively on these three issues, often from a gender perspective. I have carried out field studies on these three issues, also with a gender perspective: on the 1974 famine in Bangladesh and the 1985–7 drought in western India; widowhood in rural India; and paid work, especially of working poor women, in both India and Bangladesh. This chapter is in four sections, one for each issue and a conclusion linking all three. In sections I–III, I briefly introduce what Sen has written on each issue, summarize the findings of my research on the topic, and then highlight certain factors that serve to illuminate or complicate the issue. In the conclusion, I reflect on Sen's entitlement and capability theories and the normative concepts that they offer to those who seek gender justice in South Asia.

Before I begin, a major caveat is in order. I have read Sen selectively, not systematically. Sen has most likely addressed somewhere at some time many of the issues or questions that I will raise. But I trust that these issues or questions are worth raising nonetheless. My intent in writing this paper is to use my research to illuminate, nuance or, in some cases, problematize a related set of Sen's concepts and theories. In so doing, I also try to illustrate the links between these concepts and theories.

I should also note that my research focuses on those who live and work at the edge of subsistence. Over three decades in Bangladesh and India, I have interviewed thousands of women and hundreds of men from poor households and disadvantaged communities. Such persons often have few choices; they often face a trade-off between physical survival and other capabilities. For such persons to secure their physical survival, their labor power is often their only economic resource. Yet prevailing social norms often restrict the physical mobility of women—and thereby their ability to seek paid work—and the occupational mobility of both men and women. Further, these social norms are often reflected in official and public norms that determine the quantity and quality of paid work available to women in particular. My perspective on Amartya Sen and development issues more generally is informed by the daily struggle of women from poor households and disadvantaged communities in South Asia (and elsewhere) to seek paid work or otherwise secure a livelihood for themselves and their families.

I. Famine

I.1 Famine in Bangladesh

In 1974, Bangladesh suffered a major famine. From the end of June to early September that year, Bangladesh also suffered severe floods. Starvation began immediately after the floods, and grew in severity. The government of Bangladesh officially declared famine in late September. From early October to late November, the government operated *langarkhanas* (gruel kitchens) to provide free cooked food to destitute people (Sen 1981). At the peak of this relief effort, some 6,000 *langarkhanas* were providing free cooked food to 4.35 million people (more than 6% of the total population of the country) (Sen 1981: 131). Estimates of how many people died in the famine vary greatly, from the official estimate of 26,000 deaths to a private estimate of 80,000–100,000 deaths in one district alone (Rangpur) (ibid.). What is clear is that, without the massive relief effort, mortality would have been higher.

In Bangladesh at the time, it was widely assumed that the following series of events had taken place: severe flooding; decline in rice production; rise in rice prices; deaths by starvation. Amartya Sen did not accept this standard story-line. He reviewed data on rice production throughout 1974 and on the occupations of the destitute being fed at the *langarkhanas* in late 1974—and came up with a different story-line. According to the evidence analyzed by Sen, 1974 was a peak year in terms of both total and per capita output of rice in Bangladesh; the so-called famine districts (with high starvation mortality) had relatively high availability of food grains per capita, but wages had decreased relative to rice prices in those districts; and landless laborers, not small farmers or share-croppers, were over-represented among the destitute being fed at the *langarkhanas*.

The 1974 Bangladesh famine is one of four famines, together with the 1943 Great Bengal famine, the 1972–4 Ethiopian famine, and the 1968–73 Sahelian famine, analyzed by Sen in his 1981 book *Poverty and Famines: An Essay on Entitlement and Deprivation.* Through his empirical analysis of these historical famines, Sen found that famines have occurred even when there was no decline in or shortage of per capita food output. He therefore criticizes the then dominant view that famines are caused by a shortage in per capita food availability—what he calls the Food Availability Decline (FAD) view. Rather, Sen argues, starvation should be understood as a function of the distribution of food, of purchasing power over food, not just of food availability per se. It is important, he emphasizes, to distinguish between decline in food availability and decline in purchasing power (what he calls "entitlement to" or command over food). In Sen's entitlement theory, a person's ability to command food depends on what she owns, what she earns, what is given to her free, and what is taken away from her. What she can earn depends on what she can sell and at what price. Her income can drop sharply for a number of reasons: loss of work, failure of

production, or change in relative prices. During the 1974 Bangladesh famine, after the monsoon floods disrupted normal employment opportunities, the wages paid to landless laborers fell as the price of rice rose: hence, landless laborers (and their families) comprised the major share of those who were fed at the *langarkhanas* or died of starvation.

In late November 1974, the government of Bangladesh declared that the famine was over and closed the *langarkhanas*. But the effects of the famine dragged on. As one symptom of the lingering crisis, women continued to beg on the streets of towns and cities across the country. When the government (with international support) introduced food-for-work schemes throughout the country in late 1975 to provide additional relief, women came forward in record numbers. Women begging on the streets and seeking paid work outside their homes was uncommon in Bangladesh, where social norms have traditionally confined women to their homesteads or, if they need to seek paid work, to the homesteads of others in their villages.

In mid-1975, I began working for BRAC, today a very large and world-renowned Bangladeshi NGO, as head of its women's program. Shortly thereafter, with food aid from the World Food Programme and other agencies, the government of Bangladesh started the nationwide food-for-work program. In November 1975, on a field trip to a rural area west of Dhaka city, I met a group of women sitting under the shade of a tree near a food-for-work site. The women had just been refused work at that site. Local officials turned them back, saying: "Women in Bangladesh should not work outside their homes. We have never hired women at food-for-work sites."

Bangladesh has a long history of food-for-work programs in lean employment seasons and after floods, drought or other crises, but the participation of women was a new phenomenon. The government was simply not prepared to accept women into its rural works programs. Conditioned by tradition to believe that women should not seek paid work outside their homes, local officials regularly turned women back. Troubled by the situation, activist women friends and I began to meet and interview women informally at food-for-work sites in different parts of the country. We found that the women faced two main problems: many were denied work and those who were allowed to work were paid less than men. This was because payments were tied to the amount of earth moved and women were able to lift and carry less earth than men in a given day, as they had no tradition of doing heavy manual labor, even agricultural field work. We began to discuss these problems with responsible government officials and food aid donors.

In March 1976, the Ministry of Relief and Rehabilitation requested advice on the integration of women into its food-for-work activities. Our ad hoc group was formalized into an Advisory Committee on Women's Participation in Food-for-Work, with a mandate to advise the government and food aid donors on how to integrate women fully into the food-for-work activities. After a number of field visits and numerous interviews with women workers, the Advisory Committee met with then President Zia-ur-Rahman and his cabinet. We made two

recommendations: namely, women should be recruited in equal numbers to men and be paid the same as men (by means of a separate work-payment norm for women).

The President and his cabinet endorsed our recommendations. However, some of the food aid donors resisted the notion of a separate work norm for women, arguing that women workers were generally supplementary, not primary, earners; that women should be able to lift and carry as much earth as men in a given time period; and that separate norms would encourage false entries in the work registers (i.e. would create a perverse incentive for officials to list men workers as women workers in order to be allotted larger amounts of wheat to complete specific food-for-work schemes).

In order to be able to respond to these objections, the Advisory Committee asked two of its members to undertake a research study to determine the marital status, dependency ratio, work output, and energy requirements of women workers at food-for-work sites. With a team of investigators, we interviewed 303 women at 11 food-for-work sites in nine districts of Bangladesh. We found that women of all age groups worked at the sites, although most were between 20 and 50 years old. Only one-third were married, and many of their husbands were either unable to work or did not earn enough to feed their families; another third were widowed, and many of their sons were unwilling or unable to take care of them; and the remaining third were unmarried, deserted or divorced. All came from households with male earnings insufficient to meet subsistence needs. Just under half (47%) were the primary income earners for their families; they had, on average, 3.7 dependents each (Chen and Ghuznavi 1979, 1980).

As part of our research, we commissioned a time-motion study to investigate the work effort and output at food-for-work sites. This study found that, on average, women moved 40 cubic feet of earth per day while men moved 70 cubic feet per day. To move these respective amounts of earth, given the average female and male body weight in Bangladesh, the researchers estimated that women would require 2,100 calories and men 3,000 calories. A payment system for both sexes, based upon the male work norm, would therefore result in women receiving little more than half (57%) of the wheat payment that men received. Yet women's nutritional requirement for a day of lifting and carrying earth was estimated to be over two-thirds (70%) that of men (Chen and Ghuznavi 1979).

In sum, our study found that the women who sought food for work were from very poor households and were often the primary earners in those households, and that the work-payment norm of the food-for-work program was unfair to women. This evidence served to legitimize the Advisory Committee's policy recommendations. Meanwhile, a special food-for-work scheme for women piloted by UNICEF in northern Bangladesh had demonstrated that women were willing and able to do work for food. Since 1976, the government of Bangladesh and food aid donors have promoted women's participation in food-for-work by recruiting women more

actively, by reserving separate areas or tasks at all sites for women, by adopting separate work-payment norms for women, and by operating all-women sites. In addition, special food-for-work schemes for women that involve less heavy lifting, notably roadside tree-planting and road maintenance schemes, were introduced. Also, the World Food Programme and BRAC developed a program that used food aid to subsidize the start-up phase of poultry-rearing by women from poor households (Chen 1983, 1995). Thirty years later, these programs are still in operation. Over the past three decades, hundreds of thousands of women have participated in special women-only food-for-work schemes and the food-subsidized poultry scheme.

Our study also showed that in rural Bangladesh, where women traditionally have been barred from paid work outside their homes, there were large numbers of landless women—in addition to the (mainly male) landless laborers whom Sen wrote about—who were destitute and needed to work to survive. Our evidence also showed that the social norms which prevented women from seeking paid work outside their homes were reflected in the attitudes of local officials who denied the women the right to participate in food-for-work schemes and that a more generalized "male bread-winner" bias was reflected in the attitudes of some food aid donors and in the official norms of the food-for-work program. As noted earlier, the 1974 famine and its lingering aftermath forced thousands of women in Bangladesh to break with tradition and begin begging on the streets and seeking work at food-for-work sites. But it took specially targeted publicity and evidence to prove to the central government—and food aid donors—that the social norms being evoked by some local officials and food aid donors, and the official norms of the food-for-work program, needed to be challenged and changed if women were to be fully and fairly integrated into the food-for-work schemes.

In a male-dominated society in which women are secluded, such as Bangladesh in the 1970s, the landless rural women who sought paid work to earn food were largely invisible. In the relief effort at the height of the famine, they were considered *destitute* and therefore eligible for free food. In the relief effort after the famine, they were *not* considered *workers*, and therefore, were not considered eligible for food-for-work schemes. The fact that they remained invisible as eligible workers in the food-for-work schemes was due to the fact that official norms—local, national and international—reflected not only culture-specific norms of female seclusion but also a more generalized "male bread-winner" bias. In this context, it took informed women activists—not the free press or general public—to highlight the fact that countless invisible destitute women needed paid work to survive.

I.2 Drought in India

Since *Poverty and Famines*, in which he elaborated his entitlement theory, Amartya Sen has also made the case that famines do not occur in democratic countries

because the free media publicize scarcity or starvation, thus triggering relief programs by the government and/or private sources (Sen 1982; Drèze and Sen 1989). To illustrate his point, Sen contrasts democratic India, which has not suffered any famines since it gained independence from the British, with autocratic communist China, which suffered a major famine. The last famine in India—the 1943 Great Bengal famine—took place before independence in 1947. Since then, the government of India has been induced by the free media and a vocal public to adopt relief measures—either free food or food-for-work—to those who face a shortfall in food, either seasonally or during crises. By contrast, China experienced a devastating famine in 1959/60 in which between 16 and 30 million people died. As Sen observed, the communist party officials assigned to agriculture exaggerated agricultural output. Neither the media nor the public were free to report on starvation in remote districts (Drèze and Sen 1989). What was lacking when famine threatened China, Sen concludes, was a system of political opposition and adversarial journalism (Sen 1982; Drèze and Sen 1989).

In 1987, I studied the coping strategies of households in a village in Gujarat state, India (Chen 1991). That village and countless others in western India were experiencing the third year of a widespread drought. Although there were many signs of stress, including forced migration in search of food, work and fodder, no one in the village starved to death. Each year of the prolonged drought, the state government implemented local food-for-work schemes during the winter months. The presence of food-for-work during the lean winter months helped the most desperate households and individuals in that village avoid starvation.

During my stay in that village, I observed the quiet suffering and growing despair of a pair of widows. Kamla (the mother-in-law) and Lakshmi (her daughter-in-law) shared a single room of a two-room mud hut with Lakshmi's young children. Kamla's remaining son and his family lived next door. In normal years, following the Hindu norms of widow maintenance, that son cultivated his own land and the land that had belonged to his late father and brother, giving each of the widows their harvest share. During the 1987 drought, when there was no harvest to share, he provided no support to his sister-in-law or even his mother. As the drought wore on, Lakshmi felt she could no longer maintain Kamla on her meager salary as a part-time attendant in a government day-care center. Reluctantly, with tears in her eyes, Lakshmi asked Kamla to manage on her own. The two widows continued to live together in the same room and use the same stove but began to cook and eat separately. What appeared to be joint living, the Hindu ideal, had broken down into two, and then three, separate cooking units within a single two-room mud hut. Kamla's situation became quite dire, as she was too old and frail to engage in food-for-work. As a gesture of joint concern for Kamla and for the migrant cattle herders passing by the village in search of greener pastures for their cattle, the village elders hired Kamla to sit near the cattle trough at the edge of the village and provide drinking water to the migrant cattle herders (Chen 1991).

In September 1987, in the neighboring state of Rajasthan, a young widow named Roop Kanwar was burnt on her husband's cremation pyre. The news of her suttee spread quickly and unleashed a flood of response. Although troubled by the news, I found myself more disturbed that so much attention should be focused on one case of suttee while there was so little public concern for—or media attention on—the economic hardships experienced by countless widows during the prolonged drought in the same region at the same time.

The special predicament of Lakshmi and Kamla in a village that had been spared death by starvation illustrates the second (often overlooked) part of Sen's China–India comparison. In making the case that democracies avoid famines because their media and public can freely draw attention to starvation or deprivation, Sen also makes the important observation that democratic India has done less well than communist China in dealing with chronic deprivation. Until its recent market reforms, China had taken a large and decisive lead over India in terms of education, health and other social indicators. This was because, unlike democratic India, communist China had placed a high priority on improving basic education and health (Sen 1982). What Sen had noted in regard to chronic deprivation was being played out in the lives of Lakshmi and Kamla. The attention of the free media and general public in India tends to be drawn to dramatic events—starvation and suttee—not to the quiet deprivation that millions suffer on a daily basis.

The lack of public and media attention to the quiet desperation of widows, such as Lakshmi and Kamla, motivated me to undertake a field study of the social and economic conditions of widows around the country. Amartya Sen encouraged me and helped raise funds for the study.[1]

II. Widowhood

In several of their joint volumes on India's economic and human development, Amartya Sen and Jean Drèze have included a discussion of widowhood as a prime example of social disadvantage, gender inequality, and women's lack of agency (Dreze and Sen 1995). According to the 1990 census of India, there were about 33 million widows in India, representing 8% of the female population. The proportion of widows in the female population rises sharply with age, reaching 63% among women aged 60 and above, and close to 80% among women aged

[1] Amartya Sen raised some of the funds for this research from the World Institute for Development Economics Research (WIDER).

70 and above (Government of India 1991). As Drèze and Sen note, given this striking incidence of widowhood in India, especially among older women, the "*prospect* of widowhood reduces the quality of life of most Indian women" (Drèze and Sen 1995: 172).

Motivated by the plight of Kamla and Lakshmi and encouraged by Sen and Drèze, I undertook a study in the early 1990s of all ever-widowed women in 14 villages of India: two villages in each of seven states. Because I had observed Lakshmi and Kamla living on their own, despite the semblance of joint living, I wanted to test the notion that the joint family provides economic security to widows in rural India. Drèze had recently completed a field study of widowhood in one village in each of three states of India, so my research design was informed and guided by his analysis (Drèze 1990).

Although the circumstances of widows vary a great deal between different regions, religious communities, castes and age groups, there are certain basic factors that influence the disadvantage and insecurity experienced by many Indian widows today: strong norms of *patrilocal residence*, an important cause of the social isolation of many widows; strong traditions of *patrilineal ownership* of property, which limit the inheritance rights of widows and which modern legislation has only begun to challenge; *limited freedom to remarry*; and a *gender division of labor* and an *ideal of female seclusion* which in combination severely restrict employment opportunities for widows and other women. Around the world, a common gender division of labor places the primary responsibility for unpaid domestic and care activities on women, thus restricting the time women have to engage in paid work. Among Muslim and Hindu upper-caste communities in India, especially in northern India, the ideal of female seclusion restricts women's mobility in the public or market domains (Drèze 1990; Chen 2000).

These factors leave most widows in need of economic support from their family or community. It is widely acknowledged across India that one function of extended kinship ties is to look after the widows and children of dead relatives. The projected ideal is that Hinduism provides the widow a secure place—at least physically and economically—within the joint family. Yet, it remains unclear who exactly—which joint family or who within the joint family—is supposed to look after the widow, in what ways, or how well. The one clear norm is that daughters, once they are married into another family, should not provide support to parents, even if they are widowed and elderly (Chen 2000). The overarching question driving my research was whether widows receive economic support from their late husband's family, their parents or siblings, or their children and, if so, in what form.

What follows is a summary of my research findings on the situation of all ever-widowed women (562 in total) in 14 villages of India in the early 1990s: Bihar, Rajasthan, Uttaranchal and West Bengal in northern India; and Andhra Pradesh, Kerala, and Tamil Nadu in southern India (Chen 2000).

Residence—Of the 510 widows who did not remarry, two-thirds (67%) lived in their late husband's village, 27% in their natal village, and 6% in a third village. The vast majority (nearly 90%) lived in the same village in which they lived when their husband died: that is, strikingly few widows moved from one village to another after the death of their husband. Of those living in their natal village, only one-fourth had moved back after their husband's death. And of those living in a village other than their natal or marital village, most had been living there when their husband died.

Inheritance—Only half of the widows whose husband owned land exercised use rights over a share of that land. Of those who did, only half exercised use rights on their own: about 40% exercised use rights with sons, and the others shared rights with their daughters or other relatives. Only 9% of the widows whose father owned land exercised use rights over a share of their father's land. Not unexpectedly, widows who were married in their natal village or returned home after their husband's death were more likely than other widows to exercise use rights over their father's (or mother's) land. Finally, those widows who did inherit land often had their rights challenged or violated: those who had adult sons faced the fewest challenges while those who were childless faced the greatest challenges. In regard to housing, just over three-fourths of the widows who did not remarry lived in the same house they lived in when their husband died, and two-thirds of them owned the house.

Remarriage—Only one in ten of the ever-widowed women had remarried: 13% of the ever-widowed women in northern India, where junior levirate marriage (marrying the late husband's younger brother) is practiced, and 6% of the ever-widowed women in southern India. In both regions, lower-caste widows were more likely to remarry than upper-caste widows. Among all the ever-widowed women, 5% of Upper Caste widows remarried, 14% of Other Backward Caste widows, and 29% of Scheduled Caste widows.[2]

Gender Norms—Less than two fifths (38%) of all the widows were engaged in paid work: 8% of Upper Caste widows, 37% of the Backward Caste widows, and 56% of Scheduled Caste widows. Among those who worked outside the home, there was a decreased reliance on self-employment and increased reliance on wage employment: 6% of the widows had to give up farming and 4% had to take up manual wage employment after their husband died.

Economic Support—Less than half of the widows were maintained by their own or their husband's male kin: less than 3% lived with in-laws, less than 4% lived with

[2] Caste in India has been defined as a "closed" social group characterized by the obligation to marry within the group (Dumont 1980). The four main broad caste groups are the Brahmans, Kshatriyas, Vaishyas and Shudras. The first three are considered the Upper Castes. In the early twentieth century, the British introduced the term "Other Backward Classes" for the Shudras and the terms "Scheduled Castes" and "Scheduled Tribes" for various groups that fall outside the main caste hierarchy.

parents or brothers, and just over 40% lived with married sons. It should be noted that some of the widows who lived with married sons faced a situation similar to that of Kamla: namely, they lived adjacent to their married sons and received a harvest share but cooked and ate separately and could not rely on support when there was no harvest to share. What about the other half? They lived in what I call "female households" without adult male earners: 30% lived with unmarried children, 16% lived alone, and 4% lived with other single women (mainly other widows). Some of the last group lived with once-married daughters who had themselves been widowed, divorced or deserted.

Within this general picture, it is important to note that how these factors played themselves out in the daily lives of individual widows depended on which caste and which region of the country they were from: as a general rule, upper-caste widows, especially in northern India, face the strictest social norms and constraints (see Chen 2000 for more details).

In sum, whether widows receive economic support depends on who they live with (and where) and their bargaining power within the household which, in turn, depends on what they own and what they earn. The widow who owns land or a house—or who has an independent source of income—is more likely to be able to command support from her husband's family, her own family, or her children. However, even if she owns property and earns an income, a widow is likely to have less agency within the household than married women in that same household because of the social stigma and taboos surrounding widowhood.

What do these findings suggest about intra-household dynamics and women's agency in India? First of all, it is important to identify the religious or caste community a woman is from, the region of the country she lives in, the type and composition of household she lives in, and who is the head of the household. Second, it is important to compare her to other members of her household in terms of employment status, marital status, sex and age. Third, to gauge her bargaining power it is important to determine her fallback position. In contexts where a woman cannot depend on support from male kin, her fallback position depends on what she owns and what she earns. In sum, women's agency depends on a set of context-specific factors and understanding women's agency requires an in-depth knowledge of these factors.

My field study of widows grew into a larger project which included a workshop for widows followed by a conference on widowhood. It also generated two publications (a book on my research findings, an edited volume of the conference proceedings: Chen 2000 and Chen 1998, respectively) and a network of non-governmental organizations committed to working with widows. The three-day workshop, in which 25 widows and ten activists from the seven study states as well as Gujarat and Karnataka participated, served to place the personal experiences and common concerns of the widows at the center of the conference that followed. Their shared priorities were to have a house in their own name, a secure job or source of

livelihood, education for their children, and—last but not least—a positive social image. The workshop also stimulated interest in the creation of a network of activists and widows concerned with the problems of widowhood.

The conference that followed brought together 65 activists, scholars and policy-makers who have worked on issues relating to widows. The conference focused on four major issues: property rights, social protection, employment opportunities and the status of widows in society. For the conference, Jean Drèze and I asked P. N. Mari Bhat, a well-known demographer, to study the mortality rates of widows compared to married women. Comparing data from the 1961, 1971 and 1981 censuses of India, Bhat estimated mortality rates among widows and married women of the same older age group (above 45). He found that, on average across India, mortality rates are 85% higher among elderly widows than among married women of the same age (Bhat 1998).

Before Bhat's pioneering study, we knew that the loss of a husband often leads to a sharp decline of household income and an increase in social isolation and psychological hardship. We also knew that some widows are hounded out of their village—and even killed—by in-laws who want to reclaim the property of the late husband, and we knew that a few widows are still burned to death on their husband's funeral pyre.[3] However, until Bhat's study, we did not know the cumulative toll of social neglect of widows and social norms of widowhood in India on the life expectancy of widows. In brief, we did not fully appreciate that the consequences of widowhood for women in India are, quite simply and starkly, a matter of survival.[4]

This cumulative toll on widows has been disguised because the incidence of widowhood in India is so much higher among women (over 8% in 1991) than among men (3% in 1991). This is because the consequences of widowhood in India are very different for women and men. First, practices of patrilineal inheritance and patrilocal residence, which have the effect of dispossessing and isolating women, have the opposite effect on men. Secondly, there are no social rules dictating whether, when, or to whom a widower can remarry. As a result, there is a much higher rate of remarriage among widowed men than among widowed women, especially in the older age groups. Thirdly, men face few (if any) restrictions in seeking paid work by reason of their sex or their marital status, although their religion, caste and class may influence their employment opportunities. Fourthly, the social stigma, restrictions and taboos associated with widows do not extend to widowers.

[3] In the recent past, in many areas and social groups where suttee was once practiced—notably West Bengal—no case of suttee has been reported for several generations. However, the practice has continued in some areas—notably Rajasthan—albeit in limited numbers. Whereas only stray cases were reported throughout India from 1950 to 1980, the number of reported cases increased to as many as three a year during the 1980s, mainly in Rajasthan but also in Madhya Pradesh (Chen 2000).

[4] P. N. Mari Bhat died suddenly on 30 July 2007 at the age of 56. At the time, he was director of the International Institute for Population Sciences (IIPS) in Mumbai. Considered among the finest demographers of his generation in the world, an intellectual tribute to his work appears in the 8 September 2007 issue of the Indian journal *Economic and Political Weekly* (Das Gupta *et al.* 2007).

There are, for instance, few (if any) dress, behavior, diet or other social codes for widowers. Finally, it should be noted that men are not expected to commit suttee.

In his article entitled "More than 100 Million Women are Missing", which was published in the *New York Review of Books* in 1990, Sen focused international attention on the fact that in a large part of the developing world there are fewer females than males, not because of natural causes but because of major bias against females leading to excess female mortality (Sen 1990a). Sen compared actual female–male ratios in the populations of different countries with the ratios that would be expected if there was no gender bias in mortality (including female feticide) (Sen 1992, 2003). Since Sen's pioneering article, there have been debates about the appropriate method for arriving at exact estimates of missing women and what factors account for the missing women. The causal debate has focused on differences in the health care, medical treatment and nutrition of boys and girls, high maternal mortality rates among women of reproductive age, and, more recently with the spread of technology, sex-selective abortions. Mari Bhat's pioneering demographic analysis suggests that we need to consider excess mortality across the female life cycle, including "missing widows", particularly among older women.

III. Paid Work

Amartya Sen's doctoral dissertation at Cambridge University was on what, in its published version, he called *Choice of Techniques* (Sen 1968). In this prize-winning thesis, he made the case for choosing labor-intensive technology in poor and populous countries. His other writings on employment and technological choice are found in his monograph *Employment, Technology and Development* (Sen 1975).

More recently, Sen has drawn attention to the importance of female participation in paid work as a determinant not only of women's well-being but also of their agency, especially their bargaining power within their families. In their book *Hunger and Public Action*, Drèze and Sen point out that "there is considerable evidence that greater involvement with outside work and paid employment tend to go with less anti-female bias in intra-family distribution" (Drèze and Sen 1989: 58). They base this statement on, among other evidence, their comparative analysis of economic activity and life expectancy ratios between men and women across different regions of India (using 1980 data): the ranking of different regions in terms of economic activity and life expectancy ratios is almost the same (ibid.).

However, given the gender division of labor and the ideal of female seclusion discussed earlier, the female labor force participation rate in India is still quite low: 29% compared to the male rate of 56% (NCEUS 2007). Further, conditions of

work often limit the ability of paid work to increase women's well-being and agency (Koggel 2005). My research has largely focused on the paid work of working poor women in Bangladesh and India and has taken various forms: village studies, case studies, sub-sector analysis, policy analysis and impact assessments. More recently, with colleagues in the global research network WIEGO, I have overseen analysis of national data on employment status and employment outcomes in several countries, including India (ILO 2002; Chen *et al.* 2004, Chen *et al.* 2005). We have found that place of work and employment status are two key variables in determining the outcomes of paid work.

Each place of work—factory, firm, street, home, field, waterway, forest, construction site, or open area—is associated with specific risks, costs and opportunities. Thus, "place of work" is a key determinant of the outcomes of work, including the ability to exercise agency in the marketplace or at home. The dimensions of place of work which influence the agency outcomes of paid work include: for the *self-employed*, ownership and security of tenure of the site as well as access to customers and suppliers; for *all workers*, relationships of control at the work site (with fellow workers, with the equivalent of an employer, with other vested interest groups, with public authorities, or with family members) as well as the ability of workers to organize at the site—or away from it—in order to secure representation of their interests (Chen *et al.* 2005). Arguably, in regard to agency within the family, the key variable is whether the place of work is at home or outside the home, as interacting with fellow-workers and negotiating with vested interests in the marketplace are more likely to enhance a women's sense of identity, solidarity and bargaining power than working in isolation at home.

"Employment status" is a conceptual framework used by labor statisticians to refer to two key dimensions of the contractual arrangements under which economically active persons work, namely the allocation of the *authority* over the work situation and the outcome of the work done; and the allocation of the *economic risks* involved. As such, it is a key variable in understanding the agency inherent in different paid work arrangements. With colleagues in the WIEGO network, I have analyzed evidence from a number of countries where data are available on the composition of the labor force by employment status (Chen *et al.* 2004, 2005). Our analysis suggests a *hierarchy of earnings* across different employment statuses within the informal economy, ranked in descending order as follows: informal employers; informal employees; own-account operators; casual day laborers; and industrial outworkers.[5] Because women tend to be under-represented among informal employers and over-represented among industrial outworkers, there is a significant gender gap in average earnings within the informal workforce. Although data are not available to test this hypothesis, I would surmise that there is a similar *hierarchy*

[5] Some informal employers earn more than some formal wage workers, especially low-end wage workers in the private sector.

of agency across the different employment statuses, contributing to a similar gender gap in the degree of agency from paid work.[6]

The Government of India constituted a National Commission For Enterprises in the Unorganised Sector in September 2004. The mandate of the Commission was to examine the conditions of work and livelihood opportunities of informal workers, who represent 92% of the total workforce in India. The Commission carried out detailed analysis of the 2004/5 National Sample Survey of India and other data sources and produced its report in mid-2007. This report provides a wealth of telling data on employment rates and employment patterns across India by place of work, employment status, economic sector, and by sex.

The Commission reports that the percentage of total workers who are informally employed (i.e. without social protection) is 96% for women compared to 91% for men. What follows is a summary of the key findings of the Commission regarding women's informal employment outside of agriculture:

- The percentage of informal workers outside of agriculture who work from their own homes is 54% for women and 16% for men.
- The percentage of informal enterprises outside of agriculture located at home is 74% for women and 31% for men.
- Among informal women workers outside of agriculture, 21% are industrial outworkers/home workers, 50% are self-employed, and 29% are wage-employed.
- Among informal women wage workers outside of agriculture, just over half (53%) are regular wage workers, of which over half are domestic servants, and just under half (47%) are casual wage workers (NCEUS 2007).

Working from their own or other people's homes, whether as self-employed, industrial outworkers, or domestic workers, tends to isolate women from fellow-workers in the same trade. This isolation undermines their ability to bargain for higher wages, better working conditions, higher prices and greater market access. This isolation is particularly hard on the industrial outworkers (also called home workers), who are dependent on the subcontractor for the supply of work and raw materials and the sale of finished goods. Their lack of bargaining power has contributed to the fact that most home workers are paid very low piece rates, receive no overtime pay, often endure delayed payments, and often have their finished goods rejected arbitrarily by the contractor (ILO 2002; Chen *et al.* 2004).

In sum, whether women gain agency from paid work—and how much—depends on what they do, what employment arrangements they work under, and where they work. In his theory of entitlement, as noted earlier, Sen argues that the food a person commands depends on what she earns, which in turn depends on what she owns, what she sells, and at what price. A similar case can and should be made for the

[6] With the ILO Statistics Bureau, the WIEGO network has developed an indicator on the "structure of the labor force by employment status and sex" for measuring progress at the national level towards gender equality under Millennium Development Goal no. 3.

agency associated with different types of paid work. Admittedly, Sen focuses on paid work *outside the home* as a way of strengthening women's agency. This is an important distinction, given that the majority of women in India who are gainfully employed work from their own or others' homes and their isolation undermines their scope for agency. However, in discussing women's agency through paid work, Sen does not pay as much attention to the structure of labor market opportunities and the possibility of labor market failures as he does when discussing command over food.

IV. Gender Justice in South Asia

Most recently, Sen has made a major theoretical case that capabilities—what people are able to do and to be—provide the best basis for understanding and promoting social justice, including gender justice, or development more generally. Among other capabilities, such as political liberties and freedom of association, Sen mentions "the free choice of occupation" or being able to seek employment on an equal basis with others. But, in his theorization of capabilities, Sen stops short of endorsing a specific list of core human capabilities (Nussbaum 2005). Also, he does not fully integrate the structure of opportunities that determines whether or not persons can translate their capabilities into actual functionings.

On the basis of my research and experience in South Asia for over three decades, I would argue that gender segmentation in the structure of employment opportunities and the gender gap in access to employment opportunities are key contributors to gender inequality in well-being and social justice. This is because for poor persons, especially women, their labor power is often their only economic resource. This was true of the destitute women who were denied the right to food-for-work in Bangladesh in 1975. This was true of the widows I met in India who were denied use rights to their husband's share of land and did not receive economic support from male kin. This is true of millions of working poor women today in Bangladesh, India and other countries in South Asia.

What does Amartya Sen have to say about the structure of employment opportunities and access to employment opportunities? From my admittedly selective reading of Sen, it strikes me that his entitlement theory provides a structural perspective on opportunities while his capability theory does not (Sen 1985, 1999). Yet opportunities or the lack of them are what facilitate or constrain the ability of capabilities to become actual functionings. It also strikes me that Sen puts more emphasis on earned income and physical survival in his entitlement theory and more emphasis on non-material functionings in his capability theory. It also strikes me

that Sen puts more emphasis on groups or classes of individuals in his entitlement theory and more emphasis on individuals qua individuals in his capability theory.

To achieve gender justice in South Asia—especially for destitute women, for widows, for working poor women—will require, I believe, the promotion of labor-intensive technologies to increase employment opportunities (which, as we have seen, Sen made the case for in his Ph.D. thesis), addressing the structural constraints that limit women's access to these opportunities (as Sen argues in his entitlement theory), and increasing women's freedom and agency (as Sen calls for in his capabilities theory). It will require a focus on the basic material as well as non-material capabilities needed by women, on the structure of economic opportunities available to women, on the social and official norms that constrain certain women from accessing these opportunities, and on the collective as well as individual agency of women.

In closing, those seeking gender justice in South Asia and elsewhere remain indebted to Amartya Sen for putting famine, widowhood and paid work—and related analytical concepts and normative standards—on the mainstream development agenda.

References

Bhat, P. N. Mari (1998), "Widowhood and Mortality in India", in M. Chen (ed.), *Widows in India: Social Neglect and Public Action* (New Delhi: Sage Publications).

Chen, M. (1983), *A Quiet Revolution: Women in Transition in Rural Bangladesh* (Cambridge, Mass.: Schenkman Publishing).

—— (1991), *Coping with Seasonality and Drought in Western India* (New Delhi: Sage Publications).

—— (ed.) (1995a), *Beyond Credit: A Sub-Sector Approach to Promoting Women's Enterprises* (Ottawa: Aga Khan Foundation Canada).

—— (1995b), "A Matter of Survival: Women's Right to Employment in India and Bangladesh", in M. Nussbaum and J. Glover (eds), *Women, Culture, and Development: A Study of Human Capabilities* (Oxford: Clarendon Press).

—— (1998), *Widows in India: Social Neglect and Public Action* (New Delhi: Sage Publications).

—— (2000), *Perpetual Mourning: Widowhood in Rural India* (Delhi: Oxford University Press).

—— and Ghuznavi, R. (1979), *Women in Food-for-Work: The Bangladesh Experience* (Rome: United Nations World Food Programme).

—— —— (1980), "Rural Bangladesh Women in Food-for-Work", in A. D'Souza (ed.), *Women in Contemporary India and South Asia* (New Delhi: Manohar Publications).

—— Vanek, J., and Carr, M. (2004), *Mainstreaming Informal Employment and Gender in Poverty Reduction: A Handbook for Policy-Makers and Other Stakeholders* (London: Commonwealth Secretariat).

Chen, M., Vanek, J., Lund, F., and Heintz, J., with C. Bonner and R. Jhabvala (2005), *The Progress of the World's Women 2005: Women, Work and Poverty* (New York: UNIFEM).

Das Gupta, M., *et al.* (2007), "P. N. Mari Bhat: An Intellectual Tribute", *Economic and Political Weekly*, 8 September: 3611–15.

Drèze, J. (1990), *Widows in Rural India*, Development Economics Research Programme, STICERD, Discussion Paper 26 (London: London School of Economics).

—— and Sen A. K. (1989), *Hunger and Public Action* (Oxford: Clarendon Press).

—— —— (1995), *India: Economic Development and Social Opportunity* (Delhi: Oxford University Press).

Dumont, L. (1980), *Homo Hierarchicus: The Caste System and its Implications* (Chicago: University of Chicago Press).

Government of India (1991), *Population Census* (New Delhi: Registrar General).

ILO (International Labor Office) (2002), *Women and Men in the Informal Economy: A Statistical Picture* (Geneva: International Labor Office).

Koggel, C. M. (2005), "Globalization and Women's Paid Work: Expanding Freedom?", in B. Agarwal (ed.), *Amartya Sen's Work and Ideas: A Gender Perspective* (London: Routledge).

NCEUS (National Commission on Enterprises in the Unorganised Sector) (2007), *Report on Conditions of Work and Promotion of Livelihoods in the Unorganised Sector* (New Delhi: Dolphin Printo Graphics).

Nussbaum, M. C. (2005), "Capabilities as Fundamental Entitlements: Sen and Social Justice", in B. Agarwal (ed.), *Amartya Sen's Work and Ideas: A Gender Perspective* (London: Routledge).

Sen, A. K. (1968), *Choice of Techniques*, 3rd edn (Oxford: Basil Blackwell; Delhi: Oxford University Press).

—— (1975), *Employment, Technology, and Development* (Oxford: Clarendon Press; New York: Oxford University Press).

—— (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Clarendon Press).

—— (1982), "How is India Doing?", *New York Review of Books*, 16 December.

—— (1985), *Commodities and Capabilities* (Amsterdam: North-Holland).

—— (1990a), "More than 100 Million Women are Missing", *New York Review of Books*, 20 December.

—— (1990b), "Gender and Cooperative Conflicts", in I. Tinker (ed.), *Persistent Inequalities: Women and World Development* (New York: Oxford University Press).

—— (1992), "Missing Women", *British Medical Journal*, 304 (March): 587–8.

—— (1999), *Development as Freedom* (Oxford: Clarendon Press).

—— (2003), "Missing Women—Revisited", *British Medical Journal*, 327 (December): 1297–8.

CHAPTER 13

TIME AND INCOME

EMPIRICAL EVIDENCE ON GENDER POVERTY AND INEQUALITIES FROM A CAPABILITY PERSPECTIVE

ENRICA CHIAPPERO MARTINETTI

I. Introduction

Gender inequalities are deeply rooted all over the world and in most domains of life, and certainly gender division of labor between paid and unpaid work is a remarkable case.[1] In the last four decades women have progressively increased their

[1] Unpaid work or social reproduction activity will be understood as the total amount of activity carried out in the household context. However, at the theoretical as well as at the empirical level it could be useful to distinguish between domestic work, which includes all the impersonal activity in home management (and which could be done by people outside the family), and care work for children, the elderly or the sick, which, because of its interpersonal and emotional nature, is more difficult to hand over to others or to purchase on the market. A third category of unremunerated work is represented by the voluntary and community work supplied by civil society.

share of participation in the paid labor force but time-use surveys show that the gender distribution of unpaid work within the household has not changed in any significant way, with the undesirable (and unsustainable) corollary of a "double work day" for most women.[2]

While labor economists have directed their efforts toward understanding gender inequalities on the labor market, the consequences for well-being of an unequal, excessive burden of unpaid work on women have received less attention from mainstream economics. One of the reasons for such meager attention to this dimension of gender inequality is that, while women's participation in the labor market is visible, measurable and positively evaluated as a contribution to the market economy, unpaid work still remains largely invisible, unvalued and confined to the realm of intra-household decisions. Significant efforts have been made to include unpaid work in national and international statistics as well as to measure its economic value, but many other questions still remain open.

This chapter aims in particular to investigate how and to what extent gender inequalities in time allocation between paid and unpaid work can determine time poverty and affect income poverty and individual well-being. In section II, I will discuss at a conceptual level the reasons why the capability approach formulated by Amartya Sen (1987, 1992, 1993, 1999), which provides a broader and richer perspective for analyzing individual well-being compared to other theoretical frameworks, can play a substantial role in clarifying the nature of unpaid work and its relationship to individual and household well-being.[3] In section III, I will present some empirical evidence for such complex links between time and income in Italy, a country which unfortunately ranks high in gender disparities in the division of labor. My analysis of inequality and poverty will be conducted in both spaces using the microdata of the Bank of Italy's 2000 survey on Household Income and Wealth (SHIW).

II. Well-Being and Unpaid Work from a Capability Perspective

In the capability perspective, functionings and capabilities represent the main constitutive elements of individual well-being: functionings are achievements, the realized "beings and doings" of a person, whereas capabilities are combinations of

[2] On this, see, among others, Goldschmidt-Clermont and Pagnossin-Aligisakis (1996); Ironmonger (1996); Himmelweit (2000); McFarlane *et al.* (2000); Nyberg (2000); OECD (2000).

[3] Amartya Sen has written extensively on the profound inequalities in human capabilities which still exist between men and women in almost all countries, and many other scholars have made fundamental contributions to this field of investigation. See, in particular, Nussbaum (1995, 2000a, 2000b, 2003) and Robeyns (2000, 2001, 2003, 2008).

functionings that a person can achieve. The possibility of achieving functionings is positively related to the availability of economic resources, but it also depends on personal characteristics and external circumstances, as well as on social and cultural norms, the legal and political context, the availability of social services, access to public support and benefits, and so on. All these factors generate a wide range of interpersonal variations that greatly affect the individual process of converting resources into capabilities, and thus achieved functionings. The space of achieved functionings cannot be considered sufficient to capture the overall well-being of a person: when the purpose is to know not only what has been achieved but also what is the real freedom to achieve, then attention should be focused on the capability set rather than on a selected element of it (i.e. on the space of the functionings achieved).

The capability approach postulates a richer and more complex relationship between means and ends than in the utilitarian tradition or welfarism, and describes well-being as a process and not simply as a condition. It thus makes room for a plurality of dimensions, acts and states that are important per se (e.g. being happy and choosing what one desires). It also recognizes how important the environment (conceived in a broad sense, including household and society, culture and nature) can be in affecting human capabilities and individual well-being. All these aspects prove to be particularly relevant from a gender perspective, as I will try to show in the rest of this section.

II.1 Gender Inequalities from a Capability Perspective

The capability approach can allow us to investigate gender inequality issues in at least four distinct but related spheres: material resources, conversion factors, well-being and time allocation.

A first kind of gender inequality relates to the material basis of well-being: there is plenty of empirical evidence that testifies how resources and commodities are unequally distributed between men and women both in developed and developing countries. Everywhere women hold fewer economic resources and have lower earnings than men. These are nothing more than unavoidable consequences of persistent inequalities in the entitlement space: fewer employment opportunities, inadequate assets, limited access to credit markets, wage discrimination and lack of property rights are among the main causes of gender inequalities in the space of economic resources. To identify possible cures for this kind of inequality, it is essential to distinguish between the distribution of economic resources between men and women and the actual entitlement they have in terms of command over resources. Public policies that affect an individual's entitlement can structurally modify economic inequalities, whereas income redistributions

simply represent a partial and temporary, even if necessary, adjustment of these inequalities.

A second source of gender inequality can be noticed in the process of converting resources into functionings and capabilities. A traditional argument states that physical differences can justify differences in people's conversion rates: an ill person needs more support and help, more income and medical care, so her conversion rates will be lower than those of a healthy person. Differences in physical factors is an important key for understanding the variable ability to convert means into ends. However, if age and metabolic conditions, illness and disability, abilities and talents, are characteristics normally distributed between men and women, so that the probability of being healthy or sick is also equally distributed, there is no reason to believe that women *must* have lower conversion rates than men. Of course, women's conversion rates *can* be (and, as a matter of fact, often are) systematically lower: this happens, for instance, if a woman suffers bad health because, being a woman, she is undernourished, receives less medical care or is a victim of sexual abuse. However, the causes of this inequality must be found in cultural, social and political norms and cannot be justified only on the basis of physical features.[4] More remarkable for our purposes are, in fact, the "environmental" constraints: a woman *may* have a lower conversion rate because households, markets and societies do not offer her the same opportunities as a man. Having a lower level of education, being discriminated against in the labor market, being psychologically dependent on an adult man, or having no voice in the political arena obviously affect women's conversion ability. Thus, a clear distinction between physical features (given and often immutable) and environmental features (which, reflecting gender roles and gender norms, can be changed) is essential for appropriate public action.

A third type of gender inequality that can be measured is that of capabilities and functionings. A larger capability set will reflect greater opportunities, a higher level of freedom and a wider set of choice. A greater amount (and/or a higher quality) of achieved functionings will reflect a higher standard of living. The amount of gender inequality of this type that can be observed is, by and large, a direct consequence of gender inequalities of the previous two kinds. But once again, it can be particularly helpful in discussing gender inequality to distinguish clearly among inequalities in terms of (1) means to achieve, (2) ability to convert means into ends, (3) opportunity to achieve and (4) achievements.

From a gender perspective these distinctions are rather essential: concentrating on capabilities allows us to go back to the causes of gender inequality, whereas looking at achieved functionings gives us just a measure of the existing disparities between men and women (though an innovative measure, if compared with other standard measures based on income).

[4] For an attempt to estimate the conversion rates for subgroups of a population according to gender and age, see Chiappero Martinetti and Salardi (2008).

Finally, gender inequalities can be investigated in the realm of time allocation. Time is a resource and like any other scarce resource is unequally distributed; but it is not always easy to understand why unpaid work is so unequally distributed between men and women within the family, or to evaluate the consequences in terms of personal well-being. The empirical literature on time allocation in developed countries confirms that women do most of the unpaid work and that "everywhere husbands are scarcely available to assure a better market work/family work equilibrium for their wives" (Bonke and Kock-Weser 2001: 56). If the nature of work is different in developing countries, the situation is not substantially dissimilar: again, the overall burden of work is heavier for women than for men (UNDP 1995). Explanations for such inequalities have been found in insufficient public support, cultural norms and behavioral models that seem hard to modify, whereas socio-economic conditions such as the age and education of the spouses or the number and age of the children give only partial reasons for such differences in the patterns of male and female division of work.

If time allocation matters in determining individual well-being, how can it be encompassed within a capability perspective? If time, like income, is a (scarce) resource, should it be considered an instrument, a means for "producing" well-being? Or should it be regarded as an end and thus considered as a dimension of well-being? As in the case of income, the answer may not be obvious and might require some further clarification.

II.2 Income and Unpaid Work: Means or Ends?

It seems prima facie to be a basic tenet of the capability approach that, as already mentioned, income and, generally speaking, economic resources are not capabilities or functionings but means for acquiring commodities and services, which become inputs in securing well-being. Sen (1992, 1999) does not deny the essential contribution that income makes to achieving valuable functionings, but criticizes welfare economics for measuring poverty and inequality *exclusively* in terms of income and highlights the necessity of moving one step further by asking what people are able to do and to be with these available resources. If this distinction between means and ends is generally recognized at a conceptual level, nonetheless most empirical applications based on (or inspired by) the capability approach, including the well-known Human Development Indices, usually consider income or consumption on a par with other functionings, such as being healthy or being educated, without drawing any substantial methodological distinctions between resources and achievements.[5]

[5] A different view on the link between income and capability can be found in Bojer (2006), who argues that the capability approach should include income as a measure of access to economic goods and suggests a measure of full income capability.

Obvious justifications for this relate to the complexity of measuring functionings and capabilities, and the claim that income represents a useful proxy for capturing the material, instrumental dimension of standard of living. However, if income and economic resources should not be included in the list of capabilities, the possibility of having a job and thus earning income, and entitlement to ownership and property rights, certainly pertains to the capabilities domain. These capabilities are not only instrumentally important, as they allow people to provide means for other ends; they are also valuable per se because they allow people to achieve autonomy and self-respect, to enhance their agency and power. It is on the basis of these arguments that Alkire and Black (1997), Nussbaum (2003) and Robeyns (2003), among others, explicitly include paid work and property rights in their lists on human capabilities.

What can be said about unpaid work? Should it be considered as an instrument for achieving valuable functionings or should it be regarded as an end and be included in the list of capabilities? We need only look at our own lives to recognize that most social reproduction activities and care work provided by the household affects significantly, and in many cases contributes directly to, the generation and enhancement of our capabilities, and our possibility of achieving functionings. This is mostly true for basic dimensions of life, such as adequate nutrition and health, but generally speaking and in many particular respects, family environment is an essential source of development of individual capabilities. To receive help and care, practical, moral and psychological support, is a fundamental requirement for expressing and improving talents and abilities, generating self-respect, and living an autonomous and worthy life. Care work is not only a basic element for enhancing human capabilities but is often also an essential means of offsetting personal disadvantages and restrictions: people who have some handicap or invalidity frequently find the primary, and sometimes the only, answers to their needs within the family.[6]

Thus there seems good reason to regard domestic activities and care work as an essential input in the process of transforming commodities and resources into outputs, here conceived as a set of fundamental functionings for human life. It is, however, quite a peculiar input, not only because it is not remunerated but also because: (1) it is mainly provided by women but is principally devoted to satisfying the needs of other household members; and (2) it produces economies of scale and positive externalities, so that most domestic activities can be characterized as "public household goods" (I will return to (1) later).

The idea of seeing the well-being process as a kind of production function is not a new one: it dates back to Gary Becker's contribution to the household production literature (see Becker 1965, 1981) and it has also been mentioned in the capability

[6] Household environment can also negatively affect individual well-being. As a growing body of literature has shown, intra-household resources are not always allocated according to needs and intra-household inequality is often the rule instead of the exception in many contexts.

literature, in particular in Sen (1985), Muellbauer (1987), Kuklys (2005) and, most recently, Chiappero Martinetti and Salardi (2008) and Chiappero Martinetti *et al.* (2008).

Although there are some analogies between Becker's theoretical framework and the capability approach, there are also some substantial differences, the most relevant being their different normative foundations. Moreover, Sen (1985) raised some doubts about the appropriateness of seeing functionings as fundamental commodities produced by the household, due to the substantial and intrinsic differences between functionings and commodities. In addition, the rather restrictive assumptions of Becker and other authors regarding the existence of shadow prices and implicit markets do not really seem to fit the capability framework.

However, the capability approach, precisely because it looks beyond the commodities purchased on the market and produced by the household, can extend and complement Becker's household production theory, in at least three different directions. First, if Becker focuses on the household production function that converts goods, time and household work into useful outputs, capability theory goes further by looking at what a person can be or can do with these outputs. In this way, the capability approach allows us to link available resources (market commodities plus household commodities) with the beings and doings that a person has achieved or can achieve. Secondly, in Becker's approach the evaluative exercise focuses on a unique variable (e.g. utilities or income), whereas in the capability approach this exercise is extended to a plurality of domains of well-being. This not only places the evaluative phase in a pluralistic perspective but also permits us to analyze how and how much domestic activities and care work can affect each separate dimension of human life. Finally, unpaid work can be viewed not only in its quantitative dimension (i.e. the total amount of time spent on domestic activities), as Becker's approach does, but also in its qualitative aspects. In fact, the "value" of social reproduction activities lies not only in the amount of time devoted to that activity and in its market value, but also in terms of achieved functionings for the care-givers as well as for those who benefit from domestic and care activities.

A different way of encompasing unpaid work within the capability approach is to consider it as an end instead of (or in addition to) a means. Robeyns (2003) moves in this direction and includes social reproduction activities and non-market care, as well as time autonomy, in a list of 14 capabilities regarded as relevant from a gender perspective. In particular, she identifies social reproduction activities and non-market care as the capability of "being able to raise children and to take care of others" and time autonomy as "being able to exercise autonomy in allocating one's time". To incorporate unpaid work and time autonomy in the capability list surely gives more prominence to these aspects and can draw attention (as I also intend to do in this chapter) to the large gender inequalities that still exist in the distribution of domestic work and care responsibilities.

However, it is not wholly clear how achievements in these well-being domains should be interpreted. As Robeyns also recognizes, if, generally speaking, domestic work and non-market care might have a positive impact on those who benefit from these activities, the well-being effects on the care-givers are for various reasons much less straightforward to determine. First, as already mentioned, these activities are not equally stressful or rewarding. Second, they produce effects on a plurality of well-being domains (e.g. emotional and mental well-being, health, social relations, etc.), and all of these domains need to be considered. Finally, allocation of time among leisure, paid work, and unpaid work, particularly for women, is rarely the result of free individual choice but is strongly affected by household, society and labor market characteristics. To restrict attention only to the overall amount of unpaid work by a person and how she allocates time among different activities could give an incomplete picture and therefore misrepresent the well-being of that person.

As Sen remarks, "the issue of gender inequality is ultimately one of disparate freedoms" (Sen 1992: 125), and the distinction made by the capability framework between what people do and what people can choose to do is particularly relevant for understanding the complex links lying behind the time allocation process and the possible consequences for well-being.

A clear distinction between resources or means to achieve (e.g. amount of time), freedom to achieve (alternative time allocations among paid, unpaid and leisure activities) and achievements (amount of time effectively spent on these activities) can give a better understanding of such gender disparities. Moreover, sensitivity to the internal and external factors that might affect this conversion process (e.g. individual socio-demographic characteristics, household structures, labor market and welfare-state features, cultural norms) can throw light on asymmetries between the sexes in pursuing the life that people have reason to value and on the unequal gender distribution of resources, opportunities and achievements in different well-being domains.

In a recent paper Burchardt (2006) takes an important step in this direction, suggesting a measure of time and income poverty consistent with the capability approach. In that paper, time and income are considered essentially as input or resources. But she moves a step further, as she integrates into her analysis not only the actual time allocation between paid and unpaid work, and the consequent amount of income that can be generated, but also the range of feasible time allocations and incomes available, thus considering the capability to be free of time and income poverty. This double vision of time allocation (and, consequently, of disposable income) both as means and as ends seems particularly fruitful from a gender perspective. The analysis in the next section attempts to move in the same direction, providing some empirical findings on the two intertwined income and time spaces.

III. Gender Poverty and Inequalities in Time and Income: Some Empirical Evidence for Italy

III.1 Description of the Data

One of the main difficulties in carrying out joint analysis on time and income distribution is the lack of availability of sufficiently accurate statistical information. On the one hand, a typical data source is time use surveys, which cover time allocation in considerable detail but usually provide information on individual and household income, if at all, only in a very aggregated and rough manner. On the other hand, household surveys are generally very detailed and accurate regarding income but usually neglect time allocation. A merging of different sources is often required for carrying out this kind of analysis, with all the problems that this procedure implies (Del Boca *et al.* 2005; Burchardt 2006).

The empirical analysis presented in this section is based on the microdata of the Bank of Italy's Survey on Household Income and Wealth (SHIW), a representative survey conducted every two years on a sample of 8,000 households and over 22,000 individuals, and the main source on income and wealth distribution in Italy. The 2000 wave used for this chapter is the only one which contains an ad hoc module on unpaid work and provides information on three distinctive issues: (1) the overall amount of hours worked weekly by all household members older than 14; (2) the allocation of unpaid work among domestic care, child care, and care activities for other family members living in the household as well as for parents and relatives living outside the household; (3) the average number of hours per week contracted out for domestic work and child care or received by relatives. Finally, information is provided on family composition, employment, housing, and other individual and household socio-demographic characteristics (see Bank of Italy 2002).

The main reference unit for our analysis is individuals in work, a homogeneous subgroup of the population for which all time and income information is available ($n = 4{,}052$). Reference is made to the weekly amount of hours devoted to care and domestic activities and to paid work, while for income weekly labor income is used, which includes wages, salaries and income from self-employment recorded net of direct taxes and social security contributions. Weekly income is obtained simply by dividing annual income by 52. Most poverty and inequality index calculations are performed using the Distributive Analysis Stata Package (DASP; see Araar and Duclos 2007).

III.2 Gender Inequality in Time Allocation and Income

As already outlined in other research on this topic, Italy is traditionally considered one of the industrialized countries with the highest degree of gender inequality in terms of unremunerated work (UNDP 1995; Bonke and Kock-Weser 2001; Addabbo and Picchio 2003; Chiappero Martinetti 2003; Del Boca *et al.* 2005). A first glance at Table 13.1 confirms that there is still a substantial gender gap in unpaid work in Italy: on average women spend almost 30 hours per week in unpaid work, 2.3 times that of men (12.5 hours per week), and have an overall (paid and unpaid) workload 15% heavier than men. In our sample of over 4,000 working individuals (63% men and 37% women), almost 40% of the men spend less than five hours of unpaid work per week, while the range with the highest percentage of women (25%) is that of 21 to 30 hours of unpaid work per week.[7]

The allocation of time between paid and unpaid activities is also significantly different: men spend about three quarters of their time working on paid activities and one quarter on unpaid work, while women allocate their work time more or less equally (46% for unpaid and 54% for paid work). In absolute terms, this means that men work on average 43 hours per week on the labor market (women 8 hours fewer) but nevertheless the total workload is higher for women (64.2 hours vs 55.8 hours).

The gender gap in time allocation does not vary significantly within socio-demographic groups, confirming that gender inequality in domestic work and care responsibilities seems an intrinsic and widespread feature of Italian society. Women systematically have a cumulative workload higher than that of men, and traditional explanatory factors of inequality, such as level of education, work status or geographic area, do not seem to play a significant role. In accordance with what Gershuny (2003) identifies as a consequence of a status transition, there is a substantial increase in the amount of time people spend on unpaid work when they move from single to couple status.[8] Both men and women increase their unpaid work but women nearly treble it (from 13 to 36 hours per week) while men double it. Other personal and household characteristics do not substantially affect this pattern. No matter what level of education she has, where she lives or what her work status is, as soon as a woman moves from her single status an irreducible amount of unpaid work seems to be assigned to her in a more or less stable way that, in addition to the paid work load, confirms once again the "double work day" hypothesis, and reduces the substitution elasticity between paid and unpaid work to nearly zero, particularly for low-income families. To prove this assumption requires

[7] One man out of three does not spend any time in unpaid work at all, while this is the case for only 7% of women in our sample (figures not included in Table 13.1).

[8] We are interpreting cross-sectional data here as a proxy for life-course changes. However, these data simply compare different people of different ages and do not allow us to observe directly the transitions in the family cycle, a kind of investigation which would require longitudinal data, as in Gershuny (2003).

Table 13.1. Gender distribution in the use of time by personal and household characteristics

Unpaid work (average hours per week)				Total work (paid + unpaid) (average hours per week)			
	Female	Male	Total		Female	Male	Total
Average hours per week	29.3	12.5	18.8	Average hour per week	64.2	55.8	58.9
Range of hours per week				Range of hours per week			
≤5	12.3	39.3	29.2	≤20	1.5	1.6	1.5
6–10	10.5	19.2	15.9	21–40	12.1	20.2	17.2
11–20	16.7	25.1	22.0	41–60	37.5	50.4	45.6
21–30	24.7	8.1	14.3	61–80	31.7	19.3	24.0
31–40	15.7	4.0	8.4	81–100	11.1	5.4	7.5
>40	20.1	4.3	10.2	>100	6.1	3.1	4.2
	Female	Male	F–M ratio		Female	Male	F–M ratio
Age							
<30	18.0	7.0	2.57		55.5	49.3	1.13
31–40	33.0	14.1	2.34		67.4	58.5	1.15
41–50	31.4	15.2	2.07		65.5	57.7	1.14
51–65	35.4	11.7	3.03		69.1	55.4	1.25
Work status							
Black-collar	29.1	12.1	2.40		63.8	53.4	1.19
White-collar	29.1	13.3	2.19		62.8	53.1	1.18
Manager	22.5	14.2	1.58		62.2	56.8	1.10
Self-employed	32.6	11.7	2.79		70.8	62.5	1.13
Marital status							
Single	13.5	6.2	2.18		50.7	48.2	1.05
Married	36.4	14.7	2.48		70.2	58.5	1.20
Divorced	27.1	15.8	1.72		64.3	56.8	1.13
Widow/widower	35.9	15.2	2.36		67.8	58.4	1.16
Education							
No education	25.4	9.7	2.62		55.3	50.1	1.10
Primary school	31.8	12.7	2.50		67.1	56.6	1.19
Secondary school	27.4	12.4	2.21		63.3	55.8	1.13
University degree	29.1	12.2	2.39		61.2	53.0	1.15
Geographic area							
North	27.9	12.0	2.32		63.4	55.8	1.14
Center	28.5	12.9	2.21		63.1	56.1	1.12
South	34.2	12.9	2.65		67.5	55.5	1.22

Note: All data are for a subsample of employed or self-employed individuals ($n = 4{,}052$).
Source: Author's calculations from 2000 SHIW dataset (Bank of Italy 2002).

a more detailed analysis of the link between income and time, and I will return to this issue in section III.3.

Gender inequality is summarized in Table 13.2 and in Figures 13.1 and 13.2, where the typical tools of distributive analysis are applied to time and income distribution

Table 13.2. Inequality indices on weekly income, paid and unpaid work, and total work

	Net weekly labor income	Unpaid work	Paid work	Total work
		(hours of work per week)		
Gini index				
Overall	0.297 (0.007)	0.529 (0.007)	0.147 (0.004)	0.187 (0.003)
Men (population share = 62%)	0.293 (0.009)	0.567 (0.010)	0.126 (0.005)	0.174 (0.004)
Women (population share = 38%)	0.275 (0.009)	0.402 (0.011)	0.162 (0.006)	0.195 (0.006)
Theil index				
Overall	0.1759 (0.012)	0.5110 (0.014)	0.0453 (0.0025)	0.0599 (0.0025)
Men (population share = 62%)	0.1739 (0.017)	0.6060 (0.011)	0.0346 (0.003)	0.0523 (0.003)
Within men	0.1215 (69%)	0.2520 (49%)	0.0230 (51%)	0.0309 (52%)
Women (population share = 38%)	0.1409 (0.012)	0.2902 (0.015)	0.0524 (0.003)	0.0653 (0.004)
Within women	0.0425 (24%)	0.1690 (33%)	0.0172 (38%)	0.0267 (44%)
Between men and women	0.0120 (7%)	0.0890 (18%)	0.0047 (11%)	0.0021 (4%)

Note: n = 4,052; standard errors and percentages in parentheses.

Source: Author's calculations from 2000 SHIW data set (Bank of Italy 2002).

between men and women. Gini indexes confirm a higher concentration of unpaid work for both men and women, even if this time component is more unequally distributed in the case of men. The opposite occurs for paid work, characterized by a low Gini value but with a relatively higher concentration for women. The decomposition of Theil indexes allows us to quantify the absolute and relative contribution of the two "between-group" and "within-group" inequality components. In all three cases (paid, unpaid and total work) the within-group component plays a major role in determining total inequality, half of which is due to the "within-men" component, and from one third to 44% to the "within-women" component, while the residual can be ascribed to the disparities between genders. As expected, the between component takes a higher value in the case of unpaid work, contributing 18% to the overall inequality.

In a similar way, Figure 13.1 compares inequality between men and women in the three domains (unpaid, paid and total work), as well as in the income space, by using generalized Lorenz curves, while Figure 13.2 makes a within-group comparison using Lorenz curves.[9] The generalized Lorenz curve for women clearly dominates that of men in the case of unpaid work (Figure. 13.1a), showing that the cumulative domestic workload is systematically higher, and the distance between these two curves increases as we move to the right end of the x-axis, that is, when greater

[9] As is well known, generalized Lorenz curves allow proper comparisons in the case of distributions characterized by different absolute levels, as happens in the case of the distiribution of unpaid work between men and women, for which mean values are substantially different, while in the case of comparisons within a group, Lorenz curves are an appropriate tool.

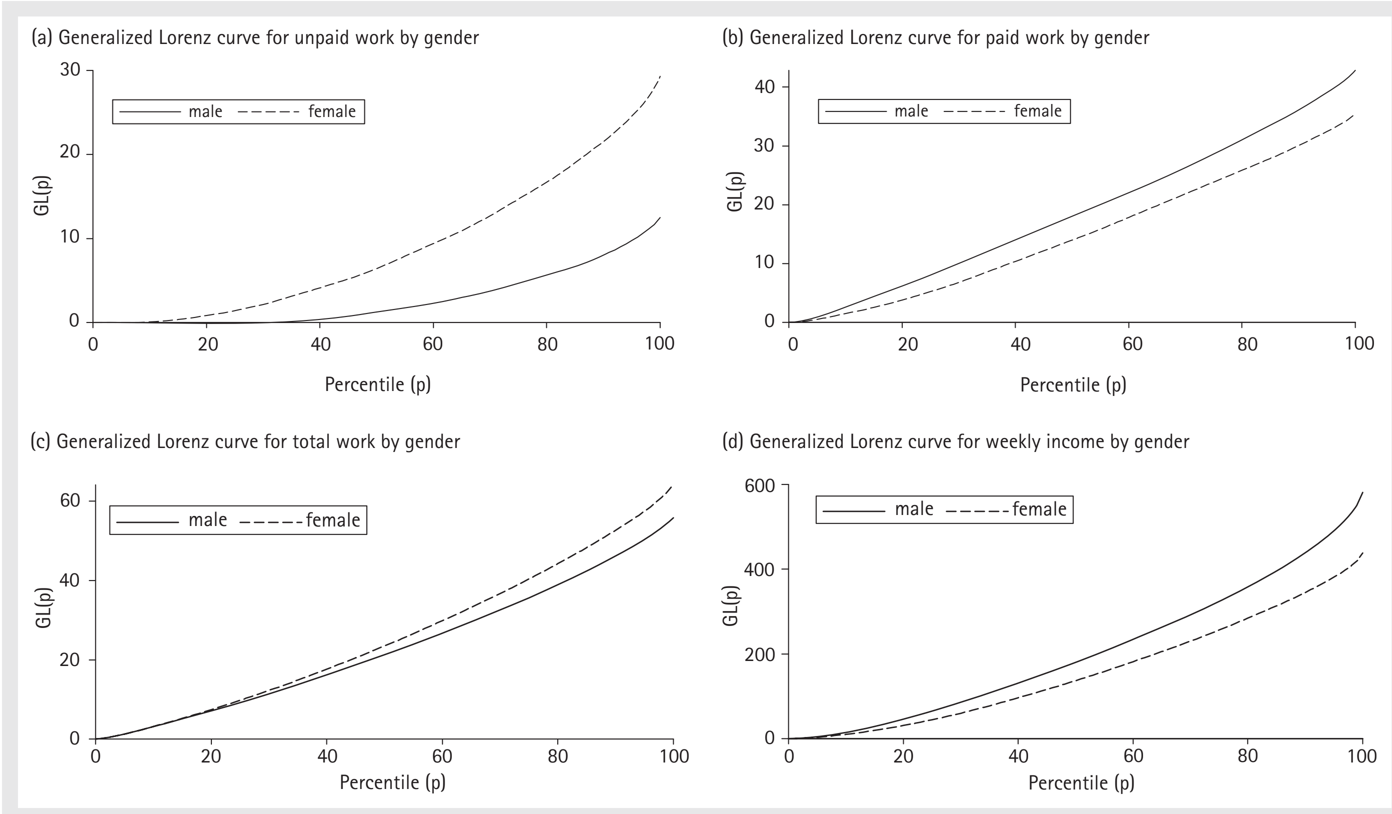


Fig. 13.1. Gender inequality between groups: generalized Lorenz curves on paid work, unpaid work, total work and weekly income, by gender

Note: $n = 4{,}052$.
Source: Author's calculations from 2000 SHIW data set (Bank of Italy 2002).

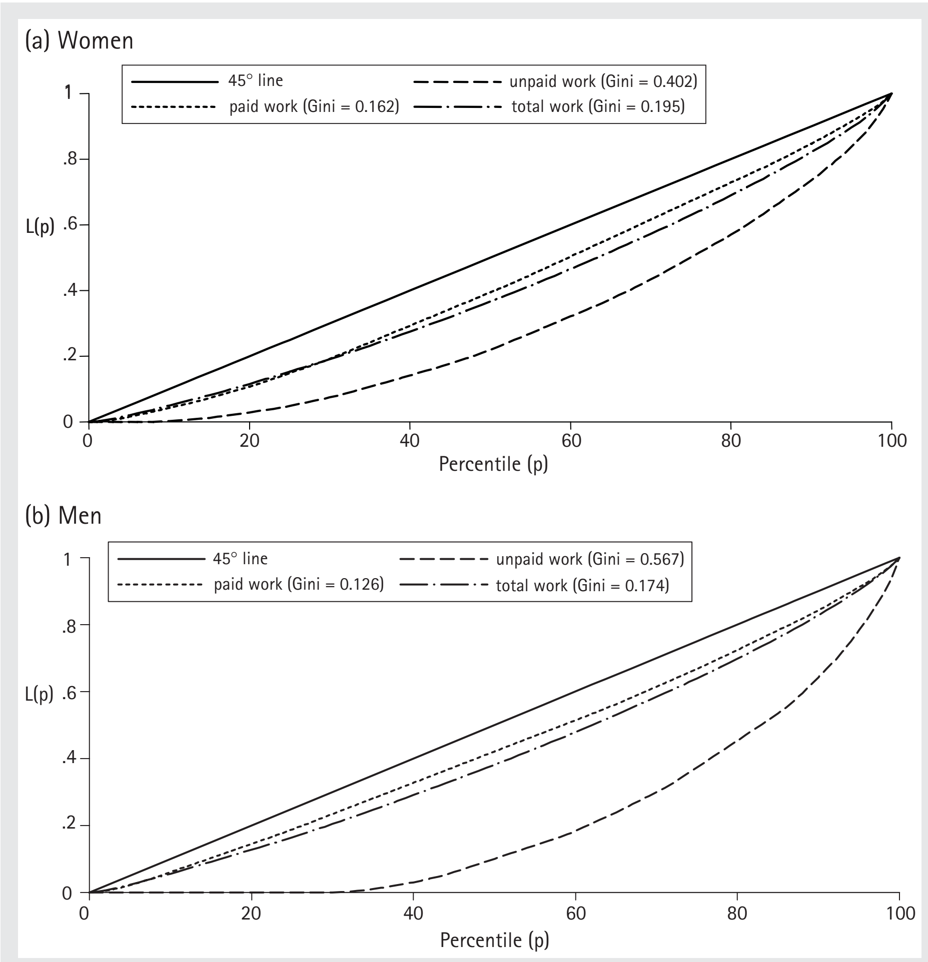


Fig. 13.2. Gender inequality within groups: Lorenz curves on paid, unpaid and total work, by gender

Note: $n = 4{,}052$.
Source: Author's calculations from 2000 SHIW data set (Bank of Italy 2002).

shares of the population are taken into consideration. The pattern is reversed in the case of paid work (Figure 13.1b) but the differences between the curves for men and women are smaller. The generalized Lorenz Curve for women is still higher when we look at total workload (Figure 13.1c).

Within-group comparisons in time allocation described by Lorenz curves separately for men and women show that unpaid work is more unequally distributed for both groups, even if more considerably in the case of men (Figure 13.2b), almost one third of whom, as already noted, do no unpaid work at all. This inequality is

partially offset by lower disparities in paid work distributions, typically characterized by full-time work for both men and women.

Table 13.2 and Figure 13.1d also indicate gender inequality in the income space, another area where horizontal inequality between groups typically adds up to vertical inequality. The Gini index for all workers is close to 0.3, slightly higher for the male subgroup compared to women, and the generalized Lorenz curve for men dominates the corresponding curve for women. However, and similar to what we see for gender inequality in time, the decomposition of the Theil index shows that most of the inequality in labor income is due to the within-subgroup component and, in particular, to inequality among men, which accounts for 69% of that inequality. The equivalent component for women gives a relative contribution of 24% to the overall inequality, while the disparities between men and women account for the remaining 7%.[10]

As we noted earlier, the capability perspective maintains that focusing on economic resources is not sufficient for evaluating individual well-being or understanding the causes of gender inequalities in income distribution. As a matter of fact, these gender inequalities generally arise from horizontal inequalities between men and women in control over resources and thus opportunities (e.g. the capability of participating in the labor market), inequalities which can arise from mechanisms of gender discrimination that act on the labor market but which in turn are also affected by gender inequalities in the distribution of domestic activities and care responsibilities.

The considerable workload of Italian women workers can help explain why Italy has one of the lowest female employment rates among the OECD countries.[11] This is partially confirmed by our data set, as almost 35% of women not lacking for work, when asked to identify the main reason, claimed that this was due to the necessity of taking care of family members, while the percentage of men not seeking work who gave the same answer is negligible (0.75). Family ties are more relevant when the woman has pre-school children (52% of these cited this as their main reason) but the percentage of childless women who do not participate in the labor market for family-care reasons is still quite high (over 30%), showing that informal care for the elderly helps to discourage female participation in the labor market (on this see also Marenzi and Pagani 2005). Finally, 11% of women do not look for a job because they wish to spend more time with their own family, while almost 10%

[10] It should be borne in mind that the relatively low level of income inequality is due to the fact that we are considering a relatively homogeneous subsample of population which includes only workers. The Gini index calculated on the equivalent income of the overall sample is higher, 0.329 (see Bank of Italy 2000). Medeiros *et al.* (2007) find different values but a similar pattern in the unpaid inequality decomposition in Bolivia.

[11] In 2000, the year of reference for our analysis, the Italian female employment rate was 39.6%, whereas the corresponding rate for the OECD as a whole was 55.3% (OECD 2004).

do not seek employment because they are discouraged by the difficulties of finding a job.

Two conclusions can be drawn from this descriptive analysis. First, these figures are given a precise significance from the capability view, in virtue of the distinction between opportunities (to have a job), preferences (to stay out of the labor force because I prefer to have more time to spend with my family) and constraints people face due to contextual factors that can restrict the real freedom to pursue their own well-being (I cannot take part in the labor market because I *must* take care of my children, and/or the labor market discriminates against women having parental responsibilities). Secondly, even if capabilities are usually considered as inherently unobservable, knowledge of people's effective or potential constraints can nevertheless contribute to an indirect estimation of the opportunity set.[12]

III.3 Time and Income Poverty

This final subsection aims to present some results of our analysis of time and income poverty, both individually and jointly considered. We will focus our attention on workers and not the whole population, so our analysis will be confined to those who are usually defined as working poor, that is individuals who have regular employment but remain in relative poverty.

Following a standard approach, the income poverty threshold is defined as €288.46 per week, corresponding to 60% of the median income, which in our sample was €480.77.[13] In a similar way, the time poverty threshold might be defined as a given percentage of median disposable time, but what constitutes disposable time is a disputed matter (Goodin *et al.* 2005; Burchardt 2006). Generally speaking, disposable time is considered to be the time left over from personal care and paid and unpaid activities. The amount of time we need to satisfy basic physiological needs is conventionally estimated at around 12 to 14 hours per day, and time use surveys using individual time diaries generally confirm these figures. In the course of a single day or week, time for paid work is also largely predetermined,[14] while it is more difficult to define a minimum or standard amount of time that should be devoted to unpaid activities.

In this chapter, we consider disposable time as the overall time available per week (168 hours) minus the amount of time spent for personal care (conventionally identified as 12 hours per day and assigned to each individual), minus the effective

[12] See Burchardt and Le Grand (2002) for an interesting account of the extent to which decisions on labor market participation are affected by constraints and opportunity.

[13] The corresponding value for the whole household is €502 per week, while the equivalent income is €270 (Bank of Italy 2002).

[14] The average time spent on paid work in our sample of workers is equal to 40.1 hours, which corresponds to a typical full-time job.

Table 13.3. Time poverty and income poverty by gender

	Men	Women	Total
Not poor in either space	70.4	44.4	60.6
Time-poor only	19.0	29.5	23.0
Income-poor only	8.5	17.9	12.0
Poor in both spaces	2.1	8.1	4.4

Notes: n = 4,502; time poverty threshold = 17.4 hours per week; income poverty threshold = €288.46 per week.

Source: Author's calculations from 2000 SHIW data set (Bank of Italy 2002).

amount of time allocated by each individual to paid and unpaid work. The median disposable time in our sample is 29 hours and the time poverty threshold, corresponding to 60% of this value, is equal to 17.4 hours per week, that is 2.5 hours per day.

As shown in Table 13.3, only 4.4% of individuals are classified as both income- and time-poor, while on average one Italian worker out of four is only time-poor and one out of eight is only income-poor, proportions that vary significantly by gender. The share of women who are time-poor is almost 30%, 10 percentage points higher than for men; the number of women below the income-poverty threshold is twice that of men, and the number of time- and income-poor women is four times that of men. The correlation coefficient between income and disposable time is −0.17 for men and −0.15 for women.

Table 13.4 shows, in turn, that on the basis of the Foster, Greer and Thorbecke class of poverty measures, the headcount ratio or incidence of poverty (P_0) indicates a significant percentage of working poor (16.5%) with a substantial gap between men and women: 11.6% of male workers, but 24.9% of female workers, are below the poverty line.[15] These large disparities between men and women are also confirmed by the other poverty measures, namely the intensity of poverty (P_1 or poverty gap) and the poverty severity index (P_2 or square poverty gap) as well as the index decomposition. Although women represent only one-third of all workers, they are the major contributing factor to the overall poverty index (56%–58%).

A similar gender pattern, but one which is worse overall, emerges for the parameter of time: more than one worker out of four is time-poor (26.7%), and although the contribution of poor men and poor women to the overall index is rather similar (48% and 52% respectively), only one man in five is time-poor, while this condition affects more than one third of woman workers.

[15] For checking the extent to which the number of poor is affected by the choice of poverty threshold we also did a sensitivity analysis. With an income-poverty threshold of 50% of median labor income, the percentages of working poor change but the gender gap remains the same: twice as many women workers are poor compared to men (20.6% vs 10%, respectively).

Table 13.4. Poverty indices in time and income space

		Unidimensional poverty		
		P_0	P_1	P_2
Income space	Overall index	16.5	6.8	3.7
Men (population share = 63%)	Subgroup index	11.6	4.7	2.5
	(Absolute contribution)	7.2	2.9	1.5
	% contribution	44%	43%	42%
Women (population share = 37%)	Subgroup index	24.9	10.5	5.8
	Absolute contribution	9.3	3.9	2.2
	% contribution	56%	57%	58%
Time space	Overall index	26.7	17.2	14.5
Men (population share = 63%)	Subgroup index	20.6	12.7	10.5
	Absolute contribution	12.9	7.9	6.6
	% contribution	48%	46%	46%
Women (population share = 37%)	Subgroup index	36.8	24.9	21.0
	Absolute contribution	13.8	9.3	7.9
	% contribution	52%	54%	54%
	Multidimensional poverty (time and income)			
		Overall	Men	Women
Chakravarty *et al.* (1998)	$\alpha = 0$	30.0	22.6	42.1
	$\alpha = 1$	15.7	11.7	22.3
	$\alpha = 2$	11.3	8.4	16.0
Bourguignon and Chakravarty (2003)	$\alpha = 1, \beta = 1, \gamma = 1$	24.1	17.2	35.5

Source: Author's calculations from 2000 SHIW data set (n = 4,052).

Finally, income and time have been taken jointly into consideration and several multidimensional poverty indexes have been computed. Table 13.4 reports

1. the Chakravarty *et al.* (1998) class of poverty measures, which represents the multidimensional extension of the FGT poverty indexes P_α; and in the simpler bivariate case it takes the form:

$$P(X, z) = \frac{1}{n}\sum_{i=1}^{n}\left[a_1\left(\frac{z_1 - x_{i,1}}{z_1}\right)^{\alpha} + a_2\left(\frac{z_2 - x_{i,2}}{z_2}\right)^{\alpha}\right].$$

   The weighting coefficients a_1 and a_2, $0 < a_i < 1$ and $\sum_{j=1}^{k} a_i = 1$, reflect the importance attached to the j dimension and α is a parameter reflecting poverty aversion.
2. the class of measures suggested by Bourguignon and Chakravarty (2002, 2003),

$$P_{\alpha,\gamma}(X, z) = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{z_1 - x_{i,1}}{z_1}\right)^{\gamma} + \beta^{\frac{\gamma}{\alpha}}\left(\frac{z_2 - x_{i,2}}{z_2}\right)^{\gamma}\right]^{\frac{\alpha}{\gamma}},$$

where $\alpha \geq 1, \gamma \geq 1$ is a substitutability parameter and $\beta > 0$ shows the relative importance attached to attribute 2 relative to attribute 1 ($\beta = a_2/a_1$). The main advantage of this index over the previous one is that it allows us to take into account substitutability and complementarity among attributes. Table 13.4 reports the simpler case in which all parameters are equal to 1, that is when the two attributes (income and time) have equal weights and are substitutes, and the iso-poverty curves are linear ($\alpha = \gamma = 1$).

All indexes show a higher incidence, intensity and severity of multidimensional poverty for women workers, with a substantial gender gap.

IV. Conclusion

The theoretical suggestions and empirical exercise presented in this chapter represent merely a first preliminary step in the direction of research that could be helpful for: (1) investigating inequalities between men and women in the interrelated domains of time and income, (2) revealing the invisible contribution of unpaid work and the impact it can have on well-being, and (3) understanding the complex relationship between constraints and preferences in time allocation decisions on the one hand, and, on the other hand, well-being in terms of opportunity to pursue the life people value.

In section II of this chapter, I discussed how these three research goals can find a proper space within a capability framework, highlighting the way in which gender inequalities can arise not only in material resources and time allocation but also in conversion factors, and ultimately, in individual well-being. Further, in discussing the distinction between means and ends with reference to both income and unpaid work, I argued that regarding domestic and care activities as an essential input into the production of well-being, i.e. converting goods and services into functionings and capabilities, means proceeding one step further from the simple amount of time allocated for these activities to the impact that these activities have on individual well-being for both care receivers and care-givers.

In the third section, I presented some empirical evidence on gender poverty and inequalities in terms of time and income in Italy, one of the most peculiar cases of gender disparity among the industrialized nations. With regard to the distribution of unpaid work, and in line with other studies, two facts clearly emerge from this analysis. First, the amount of time dedicated to social reproduction work is unequally distributed *between* men and women but equally distributed *among* women, in that it does not seem easy to explain only on the basis of factors such as social and economic status, geographical area, level of education, age, or household

size and structure. From whichever point of view the problem is considered, the burden of work is always more onerous for women, and the fact that during the last decade women have participated to a greater extent in the job market has simply translated into an overall work load—paid plus unpaid—which remains greater than men's. This double inequality in both income and time allocation cannot but be the consequence of the higher level of poverty among women workers.

The preliminary analysis presented in this chapter is only a hint of the fruitful avenues of research that the capability approach opens up for a better understanding of gender inequalities and gender poverty: much work remains to be done. With respect to unpaid work, at least two directions should be pursued.

First, our descriptive analysis provides evidence for links between job opportunities and constraints on time, links that in the case of Italy can contribute to explaining the peculiarities of this case—low female participation in the labor force, a low fertility rate, discrimination in terms of job opportunities and careers, and large gender gaps in wages. A deeper analysis could make these links more evident.

Second, the question of how an unequal distribution of unpaid work can negatively affect the well-being of care-givers in terms of professional decisions is just one side of the coin; the other is how, and to what extent, social reproduction activities supplied by women positively affect the recipient's well-being and what achieved functionings are most affected by these activities. But this is a topic for future study.

References

Addabbo, T. (2003), "Unpaid Work by Gender in Italy", in A. Picchio (ed.), *Unpaid Work and the Economy* (London and New York: Routledge).

—— and Picchio, A. (2003), "Living and Working Conditions: Perspective, Concepts and Measures", paper presented at the First Cost Action 34 Symposium on Gender and Well-Being, June, Modena.

Alkire, S., and Black, R. (1997), "A Practical Reasoning Theory of Development Ethics: Furthering the Capabilities Approach", *Journal of International Development*, 9: 263–79.

Araar, A., and Duclos, J. (2007), *DASP: Distributive Analysis Stata Package, Version 1.3* (Université de Laval, World Bank, PEP and CIRPEE).

Bank of Italy (2002), "I bilanci delle famiglie italiane nell'anno 2000", *Supplemento al Bollettino Statistico: Note metodologiche e informazioni statistiche*, 12(6) (June) (Rome: Bank of Italy).

Becker, G. (1965), "A Theory of the Allocation of Time", *Economic Journal*, 75(299): 493–517.

—— (1981), *A Treatise on the Family* (Cambridge, Mass.: Harvard University Press).

Bojer, H. (2006), *Income Capability and Child Care*, Memorandum no. 14 (Oslo: Department of Economics, University of Oslo).

Bonke, J., and Kock-Weser E. (2001), *The Welfare State and Time Allocation: Denmark, Italy, France and Sweden*, Welfare Distribution Working Paper 9 (Copenhagen: Danish National Institute of Social Research).

Bourguinon, F., and Chakravarty, S. R. (2002), *Multi-dimensional Poverty Orderings*, DELTA Working Paper 2002-22 (Paris: DELTA, Ecole Normale Supérieure).

—— —— (2003), "The Measurement of Multidimensional Poverty", *Journal of Economic Inequality*, 1: 25–49.

Burchardt, T. (2006), "Modelling the Capability to be Free of Time and Income Poverty", paper presented at the 2006 International HDCA Conference, 29 August–2 September, Groningen.

—— and Le Grand, J. (2002), *Constraint and Opportunity: Identifying Voluntary Non-Employment*, CASE Paper 55 (London: London School of Economics).

Chakravarty, S. R., Mukherjee, D., and Ranade, R. (1998). "On the Family of Subgroup and Factor Decomposable Measures of Multidimensional Poverty", *Research on Economic Inequality*, 8: 175–94.

Chiappero Martinetti, E. (2003), "Unpaid Work and Household Well-Being: A Non-Monetary Assessment", in A. Picchio (ed.), *Unpaid Work and the Economy* (London and New York: Routledge).

—— Grasso, M., and Pareglio, S. (2008), *Well-Being Assessment and Policy Simulations: Toward a Micro–Macro and Dynamic Integrated Approach Based on the Capability Approach*, HDCP-IRC Working Paper 1 (Pavia: Institute for Advanced Study).

—— and Salardi, P. (2008), *Well-Being Process and Conversion Factors: An Estimation of the Micro-Side of the Well-Being Process*, HDCP-IRC Working Paper 3 (Pavia: Institute for Advanced Study).

Del Boca, D., Locatelli, M., and Vuri, D. (2005), "Child Care Choices of Italian Households", *Review of Economics of the Household*, 3: 453–77.

Gershuny, J. (2003), *Time, Through the Lifecourse, in the Family*, Research Working Paper 2003-3 (Colchester: Institute for Social and Economic Research, University of Essex).

Goldschmidt-Clermont, L., and Pagnossin-Aligisakis, E. (1996), "Measures of Unrecorded Economics Activities in Fourteen Countries", in UNDP (ed.), *Background papers for the Human Development Report 1995* (New York: Oxford University Press).

Goodin, R., Bittman, J., and Saunders, S. (2005), "The Time-Pressure Illusion: Discretionary Time vs Free Time", *Social Indicators Research*, 73: 43–70.

Himmelweit, S. (ed.) (2000), *Inside the Household: From Labour to Care* (Basingstoke: Macmillan Press).

Ironmonger, D. (1996), "Counting Outputs, Capital Inputs and Caring Labour: Estimating Gross Household Production", *Feminist Economics*, 2(3): 37–64.

Kuklys, P. (2005), *Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications* (Berlin: Springer).

Marenzi, A., and Pagani, L. (2005), "The Impact of Elder Parents on Italian Women's Labour Market Participation", *Rivista di Politica Economica*, 34.

McFarlane, S., Beaujot, R., and Haddad, T. (2000), "Time Constraints and Relative Resources as Determinants of the Sexual Division of Domestic Work", *Canadian Journal of Sociology*, 25: 61–82.

Medeiros, M., Osorio, R. G., and Costa. J. (2007), *Gender Inequalities in Allocating Time to Paid and Unpaid Work: Evidence from Bolivia*, Working Paper 34 (Brasilia: International Poverty Centre).

MUELLBAUER, J. (1987), "Professor Sen on the Standard of Living", in A. K. Sen (ed.), *The Standard of Living* (Cambridge: Cambridge University Press).

NUSSBAUM, M. (1995), "Human Capabilities, Female Human Beings", in M. Nussbaum and J. Glover (eds), *Women, Culture and Development: A Study of Human Capabilities* (Oxford: Clarendon Press).

—— (2000a), *Women and Human Development: The Capabilities Approach* (Cambridge and New York: Cambridge University Press).

—— (2000b), "Women's Capabilities and Social Justice", *Journal of Human Development*, 1: 219–48.

—— (2003), "Capabilities as Fundamental Entitlements: Sen and Social Justice", *Feminist Economics*, 9(2–3): 33–59.

NYBERG, A. (2000), "*Power, Gender and GNP*", paper presented to the 20th EAEPE Conference, Berlin, 2–5 November.

OECD, (2004), *Employment Outlook* (Paris: OECD).

PICCHIO, A. (2003), *Unpaid Work and the Economy: A Gender Analysis of the Standards of Living* (London: Routledge).

ROBEYNS, I. (2000), *An Unworkable Idea or a Promising Alternative? Sen's Capability Approach Re-Examined*, Discussion Paper 00.30 (Leuven: Center for Economic Studies).

—— (2001), "Unpaid Work versus Jobs: A Capability Perspective", unpubl. MS, Radboud University, Nijmegen.

—— (2003), "Sen's Capability Approach and Gender Inequality: Selecting Relevant Capabilities", *Feminist Economics*, 9(2–3): 61–92.

—— (2008), "Sen's Capability Approach and Feminist Concerns", in S. Alkire, F. Comim and M. Qizilbash (eds), *The Capability Approach in Human Development: Concepts, Applications and Measurement* (Cambridge: Cambridge University Press).

SEN, A. (1985), *Commodities and Capabilities* (Amsterdam: North-Holland).

—— (1987), *The Standard of Living* (Cambridge: Cambridge University Press).

—— (992), *Inequality Re-Examined* (Oxford: Clarendon Press).

—— (1993), "Capability and Well-Being", in M. Nussbaum and A. K. Sen (eds), *The Quality of Life* (Oxford: Clarendon Press).

—— (1997), *On Economic Inequality*, expanded edn (Oxford: Clarendon Press).

—— (1999), *Development as Freedom* (New York: Knopf).

SUDGEN, R. (1993), "Welfare, Resources and Capabilities: A Review of *Inequality Re-Examined* by Amartya Sen", *Journal of Economic Literature*, 31: 1947–62.

UNDP (UNITED NATIONS DEVELOPMENT PROGRAMME) (1995), *Human Development Report* (New York: Oxford University Press).

CHAPTER 14

DEATH AND GENDER IN VICTORIAN ENGLAND

JANE HUMPHRIES
KIRSTY McNAY

I. Introduction

Amartya Sen's work on gender bias in mortality ranks among his most influential and important. In a series of papers in the late 1980s, Sen claimed that across the developing world, 100 million women were "missing": presumed dead as a result of unequal access to survival-determining resources. Sen's estimate of the number of "missing" women demonstrated that excess female deaths are no minor problem but in cumulative terms rank as one of the greatest human tragedies of our times, exceeding the combined casualties of all twentieth-century famines and dwarfing the death toll in both world wars.

Sen's pioneering contributions initiated several related research programs. The first has sought to agree the best method of computation and refine the estimates of women missing from different parts of the developing world as new and often improved census data become available (for a survey of the literature, see Klasen and Wink 2005). The second has sought to explain excess female mortality, identifying

both proximate and deeper causes. Why have girls and women died at higher rates than men and boys and what proportion of these excess deaths can be related to unequal access to resources? Why have girls and women been denied equal access to survival-enhancing resources, and what are the mechanisms of discrimination? Although the methodology for computing missing women is not uncontroversial, uncovering both the immediate and deeper causes of excess deaths has proved more difficult.

There is now a huge literature demonstrating gender bias in the allocation of life-enhancing resources such as breastfeeding, nutrition, immunization, and health care in developing countries (see e.g. Bardhan 1974; D'Souza and Chen 1980; Chen, Huq and D'Souza 1981; Rosenzweig and Shultz 1982; Das Gupta 1997). While female disadvantage constitutes a smoking gun as far as a first-level explanation of missing women is concerned, it leaves the underlying causes, the sources of disadvantage themselves, unexplained. Are girls and women prevented from accessing survival-related resources as readily as boys and men for economic reasons? Do households privilege male access to food and health care because they rely more upon boys and men for economic support? Alternatively, do cultural factors promote different child care or more ready recognition of male needs for nutrition or medicine? Are female needs hidden behind expectations that girls and women show moderation or abstention? There are also questions about the mechanisms of exclusion. How exactly does inequality occur, especially as with many of these resources, distribution takes place in the household and through processes in which women participate, indeed often control, such as serving food and visiting the clinic?

Development economists have pursued these issues using two different strategies. Where direct evidence on the gendered distribution of resources is available they have sought to correlate observed inequality with factors that might influence the economic, social and cultural worth of women and girls, such as labor market opportunities, inheritance and marriage customs, and religious practices. Absent direct evidence on the distribution of resources, the mortality outcomes themselves might be dependent on economic, social and cultural variables. The former kind of investigation often involves microstudies of households in particular communities, while the later usually involves cross-sections or even time series from national samples. In addition, the links between the economic usefulness of women and girls and the allocation of household resources have been developed theoretically through Nash-bargaining models and so put on firmer, though not uncontested, foundations.

The debate about missing women has a historical counterpart. Although studies of the past are less common, less likely to be based on extensive or firm statistical foundations, and less developed in terms of their underlying economic logic, historians have managed to show that excess female mortality was widespread in certain age groups for long periods of time (Johansson 1977, 1991; Tabutin 1978; Wall, 1981; Ginsberg and Swedlund 1986; Humphries 1991; Willems 1993; Devos

1996, 1998; Tabutin and Willems 1998; McNay, Humphries and Klasen 1998, 2005; Harris forthcoming). Moreover the evidence suggests the existence of some of the same links, from economic structure and opportunities, through social status, to disadvantage in survival, that appear in some of today's poor countries, though the historical links are perhaps marked with additional complexity and manifest in different kinds of discrimination (Devos 1998; McNay, Humphries and Klasen 1998). Urbanization, for example, might bring new ideas and erode traditional prejudices against women and girls, but if accompanied by industrialization it might involve new threats to their well-being by requiring unsuitable work in bad conditions (Tabutin and Willems 1998).

To date historical evidence suggests that higher female death rates were manifest in different age groups and therefore likely had different causes. Excess female mortality in the past was less likely than in the contemporary Third World to be a childhood phenomenon and more likely to characterize adolescence and young adulthood. Development economists have emphasized children's unequal access to resources as a major source of poor health and well-being for girls and have suggested that its remedy might lie in improved access to jobs. Anything which improves girls' and women's ability to contribute to family income is applauded as likely to enhance their valuation by other family members and increase their bargaining power within the household. In contrast, ironically, the historical evidence suggests that if girls and women become more deeply involved in economic activities with no compensating changes to their roles in childbearing, child-rearing and domestic work, their double, even triple, shifts could constitute new insults to health and well-being (Tabutin and Willems 1998). In addition, the greater likelihood of excess female mortality among teenagers and young adults in the past has led historians to implicate maternal mortality and epidemiological factors, particularly female susceptibility to tuberculosis, as key proximate causes, though whether these provide a satisfactory explanation or are quantitatively sufficient to account for excess deaths remains dubious (Humphries and McNay 1998).

Historians of excess female mortality face difficulties not encountered by development economists. National vital statistics are often not available sufficiently early for differences in mortality to be traced and linked to changing economic conditions. For the period 1600–1800, often the only data available are those reported in studies of small populations, which are generally based on the reconstitution of families from parish registers. The establishment of civil registration and the introduction of censuses permit the study of sex differences in mortality at the national level for many countries. But while this evidence is invaluable in establishing patterns and trends, the nation-state is too large and heterogeneous a unit of analysis to cast much light on the origins of excess female mortality. Without the household-level data that have been so useful to development economists, the best that historians can do is to exploit the widely cited regional and spatial differences in mortality by sex (Eggerickx and Tabutin 1994; Devos 1996, 1998; McNay, Humphries

and Klasen 2005). Unfortunately the size and heterogeneity of the geographical units for which data are available often provide only a poor basis on which to measure the effects of hypothetical explanatory variables (Tabutin and Willems 1998). Nonetheless regional analysis seems to offer the most hope to historians keen to unravel the origins of gender differences in mortality, and this is the tack we shall take here.

This chapter develops these themes in the context of Victorian England and Wales. The second section introduces the descriptive statistics of Victorian mortality, focusing on variation by age and region. Section III returns to the issue of modeling gendered death rates, thinking in terms of risk of exposure, and resistance, to disease, working these out in terms of household bargaining. Readers are warned that many of the standard explanatory links connecting the reduced value of women and girls to their lower life chances translate badly into the European past. Consistent with the advice that regional analysis offers the most promise for historians, section IV pursues an empirical investigation of male and female mortality by registration districts for mid-Victorian England and Wales. An interesting finding emerges from studying gendered death rates by age group. Although male and female death rates are systematically and predictably related to various economic, environmental and cultural factors, these differ over the life cycle. In childhood and infancy the sources of strain on males and females appear similar, but by mid-life they have diverged. Thus the same factors that threatened harm to little boys also insulted the health and well-being of little girls. Places and circumstances that were adverse for the one were adverse for the other. But as boys and girls grew into men and women, their circumstances began to take a different toll and factors to do with reproduction and the local economy began to drive wedges between their mortality outcomes. By the time they married, local factors had a different impact on men and women; what was an insult to the health and well-being of one gender was tolerated, perhaps even enjoyed, by the other. Thus the chapter concludes that excess female mortality in the past appears to be related in a more complex way to the economic and social setting, and explanations of it need to take account of the gendered division of labor in both production and reproduction.

II. Excess Female Mortality in Victorian England and Wales

At first glance England and Wales seems an unpromising area for the study of excess female mortality. National evidence on relative mortality rates, available

from 1837 when vital registration began, seems to discredit the idea of gender bias. Indeed the third English life table (3rd ELT) for 1838–54 shows English women to have an overall mortality advantage (Farr 1885; see also Tabutin and Willems 1998: annex table A.1). However, disaggregation by age group reveals a less rosy picture. From the beginning of vital registration, age-specific death rates show that aggregate female mortality advantage was confined to infancy, early childhood and older age. At other ages, where in other populations females have done better than males, they died at higher rates. Thus between the ages of 10 and 20 about 5% more females than males died, and between 30 and 40 female mortality was about 10% higher (Mitchell and Deane 1962: table 13). Moreover, evidence from the era before vital registration compiled by Wrigley *et al.* (1997) suggests that more severe female mortality disadvantage, affecting an even wider age range, characterized the later eighteenth and early nineteenth centuries (Humphries and McNay 1998; see also Harris forthcoming). Nor did differences fade out entirely after the mid-nineteenth century. In particular, girls in their teens continued to die at higher rates throughout the nineteenth century, and as late as the 1911/12 English life table female mortality exceeded male mortality at all ages between 9 and 15 (Anderson 1990; McNay, Humphries and Klasen 2005). As well as variation by age group, regional differences in relative female mortality are also masked by the aggregate evidence (Woods and Shelton 1997; McNay, Humphries and Klasen 2005). Most importantly, excess female mortality was essentially a rural phenomenon, although it did not always characterize rural localities. Mining areas and regions with large unskilled working-class populations were also at risk. Michael Anderson (1990) suggests that such geographic patterns in relative mortality might help unlock its economic and social causes, a theme which is taken up below.

III. Historicizing Explanations of Excess Female Mortality

Mortality depends on two factors: risk of exposure and resistance to disease. Risk of exposure in turn depends on the type and intensity of contacts with micro-organisms, parasites or toxins that cause disease and that have various modes of transmission. These in turn depend on housing conditions, type of work, personal cleanliness and so on. Resistance is partly genetically determined but also depends on nutritional status and other specific factors such as access to medical care

(Johansson 1991). As Tabutin and Willems note, theories of differential mortality between the sexes usually "work backwards from age-specific death rates which are initially explained in terms of differential patterns of exposure and resistance within a specific disease environment. Differences in exposure or resistance are interpreted as a result of cultural norms regarding conduct or welfare (including both positive and negative rights), except where behavioural differences can be explicitly linked to inborn genetic differences that are physiologically expressed . . . in the same way irrespective of time and place" (1998: 47).

For economists, bargaining approaches to intra-household resource distribution provide the theoretical links between excess female mortality and economic and contextual factors by highlighting how the latter give rise to both "positive" and "negative" rights. In bargaining models households pool resources obtained from a variety of sources including wage labor, self-provisioning and income transfers. Access to these resources influences exposure and resistance. Although early models had all of the resources allocated by an altruistic household head, more recent accounts posit heterogeneous preferences among household members and allow for status and authority to influence allocation. What determines an individual's clout ("threat point") in family decision-making? For most models, it is a family member's next-best alternative to the current arrangements. Thus actual contributions to the household, representing "threat points", determine bargaining power. Testable predictions follow. For example, an increase in women's wages will, *ceteris paribus*, lead to larger improvements in a working wife's welfare than an equivalent increase in her husband's wages, since the former improves her bargaining power and by implication her share of the (expanded) resources.

Once more Amartya Sen moved the discussion forward by suggesting that worthiness and legitimacy matter in the household allocation, and that these in turn depend on commonly held beliefs about the importance of different kinds of contribution (Sen 1990). For example, labor remunerated in cash may be more highly valued than unpaid household labor and this difference may be reflected in the perceived legitimacy of claims on resources. Agarwal (1997) stresses the importance of overarching social norms in legitimizing claims. Women may be viewed as secondary earners only, even if their contributions are vital to family well-being: a perception which reduces the advantage to them as individuals of undertaking paid work. If these new themes are taken seriously, it becomes more difficult to build testable models. The links from relative contributions to welfare outcomes through bargaining power are influenced by perceptions of merit formed by values and beliefs which probably originate outside the family, and are nurtured in community and class.

But even if bargaining models are embraced as providing a useful framework, at least in some heuristic sense, within which to study the allocation of resources and

its effects, it is not clear how relevant they are in the historical context. Consider household resources in nineteenth-century England. Were life chances likely to be affected by their allocation?

Breastfeeding was certainly a necessary input into infants' survival chances, but there is little evidence of aggregate excess female mortality among babies in industrializing England and Wales (McNay, Humphries and Klasen 1998). Direct historical evidence does not suggest that mothers breastfed boy babies more assiduously or for longer than their little sisters, though this would be difficult to document (Harris forthcoming). However, the fact of excess female mortality among adult women suggests that resources flowed in a different direction. Mothers who nursed children of either sex, without nutritional compensation, may have improved the life chances of their offspring at the expense of their own. Much depends here on the perception of nursing (and perhaps mothering more generally) as having claims on nutritional resources.

Health care, now an important input into resistance to disease, was less relevant in nineteenth-century conditions, when medical services were much less effective and gender differences in access to them less likely to influence life chances. Historically it was more likely that specific gender roles associated with health care, and specifically the feminization of nursing within the family, put girls and women (particularly mothers) at risk (Anderson 1990).

Food allocation, on the other hand, was probably always important. *Ceteris paribus*, a better-nourished person is better able to resist a microbial infection. Many big nineteenth-century killers, including those which historians implicate in the *surmortalité* of teenage girls, such as tuberculosis, are associated with underfeeding (Bryder 1988). Poor nutrition could even be partially responsible for some maternal deaths, not all of which should be seen as "natural" and unrelated to discrimination. Food allocation becomes even more suspect when thought of not simply as sharing out calories but also rationing particular kinds of food. Nineteenth-century families may have rated the elimination of hunger for all family members as their top priority but allocated the available protein foods with less solidarity. Historical accounts of family mealtimes certainly suggest considerable inequity in allocation of meat and other protein foods (Oren 1974; Oddy 1976; Ross 1982). The principles guiding distribution were more to do with labor market status than sex. But if wage workers were more likely to be male, criteria relating to paid employment would be tantamount to gender discrimination. As Sen and Agarwal warn, not all work may have been viewed as equally deserving of nutritional reward. In particular, domestic work, while onerous and exhausting, was undervalued. Reproductive work, as suggested above, might not have been perceived as requiring nutritional compensation, leaving mothers again disadvantaged. Heavy physical work for wages appears to have been the nineteenth-century archetype of work which needed nutritional support. Moreover, it was

not considered ladylike to eat too well. Feminine gentility required abstention. Given gender divisions of labor, such ideas easily produced gender bias in food allocation.

In addition, the focus of bargaining models may be too narrow. The main hypothesis that follows from the conventional model—that paid work, by legitimizing and strengthening the rights of women and girls, improves their access to resources, and so their resistance to disease—may need qualification in historical applications, for paid work also has direct effects on exposure and resistance. The early industrial economy was still driven by muscle and sweat. Jobs were physically draining and workplaces unhealthy. Perhaps no marginal reallocation of food could compensate for the output of effort. Moreover if female earning was also accompanied by an increase in consumption of commodities symbolizing economic power, like alcohol and tobacco, or even conspicuous consumption of fashion and finery, then it may ironically have been harmful to women's health. Maybe Victorian women were physically ill-suited to many nineteenth-century jobs and the lifestyles that went with them (Burnette 1997). Perhaps most stress involved the combination of marriage and children with a factory job, which required long hours and regular attendance. Dependence on men may have had its costs. But if the withdrawal of women and children from the labor market was accompanied by an increase in male wages, if the alternative employment involved significant physical strain, and if women's time in the home could be used productively to raise family living standards, dependence had compensations too (Humphries 1977; Bourke 1994).

In addition, the different timing of excess female mortality in the past casts further doubt on explanations favored by development economists and worked out in household bargaining. It its modern form, excess female mortality exists among infants and small children and is explained by girls' low rates of return to their families of origin, which discourages investment in their survival. Its appearance historically among older children is problematical. The tardy operation of the social and familial mechanisms which cut off such girls runs counter to one of the basic predictions of the socio-biological approach: that risk of abuse declines with a child's age (Daly and Wilson 1988). Nor do the sources of girls' low value which hinge on inheritance practices, exogamy and inter-generational transfers survive historical transplant. The standard explanation of the avoidable death of high-parity daughters in South-East Asia in terms of the old-age security provided by sons is a non-starter in Victorian England, where elderly parents turned neither to sons nor daughters but to poor relief (McNay, Humphries and Klassen 1998). When excess female mortality is detected among adults, and often the "servers and providers" who actually allocated family resources, explanations have to look more closely at the role of the mother in family and community, and the nature of her claims upon resources.

IV. The Mortality of Males and Females by Registration District in Victorian England and Wales

In the absence of suitable microeconomic data, the relationships between gender differences in mortality and economic and social variables are investigated at the level of the registration district. Registration districts are geographical units less aggregated than counties but only admit analysis at the ecological level. While no firm inferences about the behavior of individuals can be drawn, ecological associations are of value in describing the properties of social areas or groups of individuals. Although this might seem problematical given that the underlying behavioral model relates to individuals and households, the more recent themes introduced by Sen and Agarwal concern the properties of communities and social groupings. Perhaps the theoretical and apparent historical role of social norms and shared values in determining household allocations makes a virtue of the necessity of using districts as the units of analysis.

The Decennial Supplement to the 25th Annual Report of the Registrar General is the source of the mortality data. The Supplement gives mean death rates for 1851–60 by age and sex. For London the death rates are corrected for deaths in London hospitals and for deaths in workhouses situated outside the districts to which they belong. The economic and social context of the registration districts' mortality pattern is modeled using data from the Registrar General's Annual Report for 1861 and from the 1861 Census Volumes. The sources provide data on economic structure, labor market, urbanization, demographics and housing for 626 registration districts. (For further detail on the data handling and assistance with sources see McNay, Humphries and Klasen 1998.)

Male and female death rates are modeled as a function of a set of environmental and behavioral variables which proxy for exposure and resistance. The model is estimated using regression analysis. There are three related sets of questions. First, what variables explain mortality by registration district and do they differ by gender? Second, do the same variables appear to be important in the determination of mortality in all age groups or are there differences, and if so, what do these differences say about the origins of differences in mortality? Third, are male and female death rates similarly related to the same set of independent variables or are there differences in the relationships identified, and if so, how are these differences explained? Answers to these questions involve comparing both the F-tests for loss of fit and the size and significance of the regression coefficients across the equations for the different age groups.

In general terms the model relates mortality for each age group to a set of social and economic variables. Thus:

$$M_{is} = f(X_i),$$

where M is the crude death rate for each registration district, i (1–626), and for each sex, s (1, 2). These are subdivided into groups of demographic, environmental and economic variables and defined in Table 14.1.

The predictions of the effects of the exogenous variables reflect the differences and similarities in the activities and lifestyles of males and females at different ages that have been uncovered by historians and discussed above. Three points should be noted. First, the predictions for childhood are based on the assumption (well founded in the relevant historical literature) that nurture is unbiased. Victorian parents are not thought to have suffered from either son preference or "rational" discrimination originating in the greater economic contribution assumed forthcoming from older sons compared to older daughters. Hence the sources of variation in mortality for these age groups are hypothesized as gender-neutral. Second, the effects of the environmental and economic variables become differentiated by gender in the teenage years and on into mid-life, when other sex-specific effects associated with men's and women's different roles in reproduction and family support kick in. Thus an agricultural registration district is assumed salubrious for children regardless of their gender, but for teenagers its positive effects are hypothesized as limited to boys, while girls fall prey to the adverse effects of agricultural specialization traced in many historical studies and probably associated in the British case with the gendered role of deaths from tuberculosis, as emphasized by Wrigley *et al.* (1997). Similarly, higher fertility is assumed to have an adverse impact on the life chances of little boys and girls by swelling the numbers of infants in families and distracting mothers, but in later life it particularly affects women, who face a higher risk of maternal mortality and a greater strain on health from more frequent pregnancies. Third, even if the direction of the effect on mortality is predicted to be the same for both males and females, its size and significance may be very different. For example, less cramped housing might have a positive effect on the life chances of both males and females, but have a much larger and more significant effect on females simply because they spend more time indoors.

The results of the regression of childhood mortality rates are reported in Table 14.2. For children under 5, the model performs well, explaining more than 90% of the variance in death rates for both little boys and little girls. The variables have the effects predicted. The signs are the same for both sexes for all variables and, except for the constant terms, the same variables are significant in both equations. Cultural and environmental variables both play a role in influencing death rates. The employment structure is not important, the percentage of total employment in agriculture acting in this age group as an environmental factor and more rural environments proving salubrious for both sexes. One interesting

Table 14.1. Variable names, definitions and predicted effects on male and female mortality

Names (acronyms)	Definitions	Predicted effects							
		Children		Teenagers		Young adults		Adults	
		Male	Female	Male	Female	Male	Female	Male	Female
Demographic variables									
General Fertility Rate (GFR)	births/female population aged 20–44 × 1,000	+	+	0	0	0	+	0	+
Infant Mortality Rate (IMR)	infant deaths/births × 1,000	+	+	0	0	0	+	0	+
Young Marriage (YM)	[(total marriages − brides aged <21)/(total marriages − grooms aged <21)] × 1,000[a]	0	0	0	0/−	0	−	0	−
Environmental variables									
Housing Space (Space)	(inhabited dwellings/population) × 1,000	−	−	0/−	−	0/−	−	0/−	−
Logarithm of Population Density (LnPD)	log (population/size of registration district)	+	+	+	+	0/+	+	0/+	+
Mean Population (MnP)	average population, 1851–60	+	+	+	+	+	+	+	+
Cultural variable									
Female Literacy Rate (Lit)	(total marriages − brides signing with X) × 1000	−	−	−	−	−	−	−	−
Economic variables									
Per Cent Agricultural Employment (Ag)	percentage of employed men and women aged 20+ who were employed in agriculture	−	−	−	+	−	+	−	+
Per Cent Working-Class Employment (Prol)	percentage of employed men and women aged 20+ who were employed in working-class non-agricultural occupations (trade, heavy industry, and manufacturing)[b]	0	0	0	0	−	+	−	+

[a] YM captures the age difference between marriage partners. Lower YM values mean that more women are marrying underage relative to men, which may reflect limited economic opportunities and act as a mortality risk factor by weakening their bargaining power in households (Klasen 1998).

[b] The control group is the percentage of total employment in middle-class jobs (domestic service, public service, property ownership, gentry and commercial service; see McNay, Humphries and Klasen 2005).

Table 14.2. Determinants of the mortality of male and female children

Explanatory variables	Age groups			
	Under 5		5–9	
	Males	Females	Males	Females
Constant	9.964	12.439**	9.859**	8.809
GFR	.120**	.105**	.000	.001
IMR	320.525**	255.506**	10.697**	11.242**
YM	4.626	2.460	.623	1.821*
Space	−33.486**	−26.598**	−15.214**	−10.958**
LnPD	3.183**	3.170**	.204**	.147**
MnP	.048**	.058**	.008**	.010**
Lit	−11.988**	−11.949**	−2.933**	−3.337**
Ag	−7.416**	−6.733**	.154	−.211
Prol	5.568	4.891	1.195	−.435
R^2 (adjusted)	.906	.906	.469	.371

* Significant at 90%; ** significant at 95%.

finding is the replication in this historical data of a correlation found to be of importance in many developing countries: the life-preserving influence of female education. Where women's literacy rates were relatively high, both male and female children benefited. The only gender difference of note is the much larger coefficient of infant mortality rates in the equation for boys, which is readily explained by male infants' physiologically greater frailty. For children aged between 5 and 10, the model performs less well, though it is still able to account for over one third of the variation in death rates by registration district (smaller for this age group) for both boys and girls. Although the deleterious effect of fertility on the life chances of both boys and girls has faded out for this age group, the environmental variables have the same (predictable) effects as for younger children. Again there is little difference in the factors seen to influence the death rates of either boys or girls; the signs on the explanatory variables are the same in almost all cases and the same variables prove significant in both equations. The only exceptions are the economic variables, and these are not significant in either equation.

Table 14.3 reports the results for adolescents. For the age group 10–14 the model does not perform well, with girls' mortality particularly poorly explained, but what is clear is that by the time boys and girls were into their teens the differences in their experiences and lifestyles had begun to register in the processes associated with elevated or reduced death rates. The signs and significance of coefficients are no longer the same in both equations. Thus infant mortality has ceased to be positively associated with male death rates but maintains its link with female death rates, suggesting a greater influence of local environment on girls' life

Table 14.3. Determinants of the mortality of male and female teenagers

Explanatory variables	Age groups			
	10–14		15–19	
	Males	Females	Males	Females
Constant	7.462**	4.145**	8.342	4.211**
GFR	.002	.000	.002	.003
IMR	2.344	4.389**	2.041	10.761**
YM	−.770	.018	2.037*	−2.644**
Space	−4.459**	5.990**	1.154	15.303**
LnPD	.141**	−.039	−.052*	−.123
MnP	.004**	.000	.004	−.001
Lit	−2.922**	−1.270**	−4.609**	−1.095*
Ag	−1.171*	.608	−4.151**	2.281**
Prol	.633	−.329	−.927	1.682*
R^2 (adjusted)	.265	.097	.250	.248

* Significant at 90%; ** significant at 95%.

chances, operating presumably through their greater seclusion. More puzzling is the positive and significant coefficient of the Space variable in comparison with its negative and significant coefficient in the equation for males aged 10–14 and in all equations for younger children of either sex. It might be that this independent variable is particularly poorly specified, as it makes no allowance for the size or quality of housing. Authors have shown that above-average house sizes in middle-class registration districts meant that the larger households also found in these areas were accommodated without discomfort (Bristow 1985). Alternatively, perhaps more space burdened women and girls with heavier domestic labor. For young men and women aged 15 through 20 the differences widen. The gap in the size of the coefficient on Infant Mortality Rate (IMR) widens and the variable is significant in the case of females but not males. Interestingly Young Marriage (YM) becomes significant in this age group as predicted, affecting females negatively and males positively. The more frequent is female underage marriage, the higher the local death rate for young women, either because of their poor bargaining position *vis-à-vis* older husbands or more likely because of the youth of primagravidas.

Table 14.4 turns to young adults. For young adults, gendered differences in reproductive burdens and in employment differentiated the causes of mortality. For 20- to 24-year-olds fertility was positively associated with female death rates, boosting maternal mortality and adversely affecting health more generally, but it had no effect on male mortality. Infant mortality rates were negatively associated

Table 14.4. Determinants of the mortality of male and female young adults

Explanatory variables	Age groups			
	20–24		25–34	
	Males	Females	Males	Females
Constant	16.466**	1.876	17.570	7.672**
GFR	−.004	.019**	−.003	.004
IMR	−13.239**	7.793**	−9.986**	8.275**
YM	3.353**	−2.140	3.651**	−2.651**
Space	5.560	15.490**	.091	2.803
LnPD	.004	.065	.290**	−.056
MnP	.000	−.004	−.004	.000
Lit	−7.067**	−.151	−5.071**	−.500
Ag	−6.686**	4.626**	−8.773**	3.137**
Prol	−5.869**	2.333**	−8.069**	3.990**
R^2 (adjusted)	.118	.284	.236	.165

**Significant at 95%.

with male death rates (perhaps because of a reduced dependency burden) but positively associated with female death rates, as children's and women's health risks are interrelated through pregnancy and childbirth. Young marriage, as in the previous age group, had a negative impact on female life chances but was positively associated with those of men. Space also has the same strange and differentiated effect on male and female death rates. But it is in this age group, by which time men and women have accumulated work experience, with all that this involved in early Victorian England and Wales, that the effects of the economic factors begin to diverge. Agricultural specialization lowered male death rates, perhaps because it involved more salubrious working conditions, but it raised female death rates, a relationship commonly found in historical studies, perhaps because of the reduced ability that rural women had to contribute to their families' income and the lower esteem in which they were held. Similarly, more manual local economies appeared to benefit men, compared with an employment structure dominated by middle-class jobs. But the opposite was true for women in this age group. Together these effects suggest that middle-class men experienced relatively unhealthy lives compared with their working-class peers, a difference perhaps rooted in the consumption habits of the former. On the other hand, middle-class women enjoyed relatively healthy lives compared with their working-class counterparts. The same effects are manifest in the analysis of men and women aged 25 to 34. Although the positive effect of higher fertility on women's death rates is no longer significant, IMR and YM have the same (significant) effects on both male and female mortality. Similarly, local economic

Table 14.5. Determinants of the mortality of male and female adults

Explanatory variables	Age groups			
	35–44		45–54	
	Males	Females	Males	Females
Constant	26.441**	11.469	28.840**	16.907**
GFR	−.012	.002	−.011	−.009
IMR	−.670	12.624**	3.925	9.081**
YM	−1.837	−1.618	.408	−.807
Space	−17.158**	−4.684	−32.694**	−20.629**
LnPD	.586**	.137	.859**	.489**
MnP	.001	−.003	.011**	.013**
Lit	−3.276**	−.626	−2.520**	−.807
Ag	−9.985**	−.135	−9.811**	1.108
Prol	−7.278**	1.802*	−2.799	3.930**
R^2 (adjusted)	.500	.284	.638	.492

* Significant at 90%; ** significant at 95%.

structure continues to have a different impact on men's and women's death rates.

Finally, Table 14.5 reports results for men and women in mid-life. For older adults the model is more explanatory. The environmental variables become important again in these age groups, with the signs as predicted and always significant for those aged 45 to 54. Infant Mortality Rate remains positively and significantly associated with women's death rates, a continuing reminder of the costs they paid in bearing and nursing children. The dominance of agricultural employment continued to favor men but its adverse effects on women petered out for these older adults.

Thus the regressions for males and females of different ages strongly suggest that, whereas in childhood the same variables that boded ill for little boys also harmed little girls, as these children grew up and became adults, their experiences became increasingly gendered. Female life chances were significantly affected by local demographic regimes and by the prevalence of underage marriage. Economic variables, unimportant in younger age groups, also differentiated experience: for example, an agricultural setting appears to have been favorable for men but unfavorable for women. Gender discrimination was bound up with men's and women's different roles, and so mortality experiences diverged as a cohort aged and as gender moulded the way in which individuals interacted with their environments.

Table 14.6. Tests for structural difference

	Males	Females	Semi-pooled
Under 5 years of age			
Sum of squares	17,085.507	13,673.479	30,799.697
F-statistic (9, 1,232)	0.163 < 1.88, so hypothesis that coefficient vectors are the same is not rejected		
	Males	**Females**	**Pooled**
Under 5 years of age			
Sum of squares	17,085.507	13,673.479	56,589.647
F-statistic (10, 1,232)	103.459 > 1.83, so hypothesis that coefficient vectors are the same is rejected		
5–9 years of age			
Sum of squares	1,028.739	1,014.631	2,072.983
F-statistic (10, 1,232)	1.785 < 1.83, so hypothesis that coefficient vectors are the same is not rejected		
10–14 years of age			
Sum of squares	565.725	633.905	1,430.033
F-statistic (10, 1,232)	23.655 > 1.83, so hypothesis that coefficient vectors are the same is rejected		
15–19 years of age			
Sum of squares	1,218.010	1,374.096	3,609.289
F-statistic (10, 1,232)	48.345 > 1.83, so hypothesis that coefficient vectors are the same is rejected		
20–24 years of age			
Sum of squares	2,835.872	1,776.905	5,419.114
F-statistic (10, 1,232)	21.54 > 1.83, so hypothesis that coefficient vectors are the same is rejected		
25–35 years of age			
Sum of squares	2,359.353	1,287.752	4,646.068
F-statistic (10, 1,232)	33.749 > 1.83, so hypothesis that coefficient vectors are the same is rejected		
35–45 years of age			
Sum of squares	3,227.471	1,630.269	5,911.930
F-statistic (10, 1,232)	26.736 > 1.83, so hypothesis that coefficient vectors are the same is rejected		
45+ years of age			
Sum of squares	4,812.322	2,434.290	9,447.974
F-statistic (10, 1,232)	37.425 > 1.83, so hypothesis that coefficient vectors are the same is rejected		

Note: f is such that Prob [F n_1, $n_2 \leq f$] = .95.

This hypothesis can be formally tested by comparing the unrestricted results of the regressions relating to the separate equations for males and females with the result for a pooled regression where the coefficients are restricted to be the same for

males and females. An F-test can then be used to test for structural difference by gender. Table 14.6 presents the relevant statistics for the unrestricted and restricted equations by age group.

For all age groups except 5- to 9-year-olds the hypothesis that the coefficient vectors are the same is rejected, that is, the explanatory variables have different effects on the mortality of males and females. But the analysis needs qualification in the case of children aged less than 5, where the main difference is the significantly larger positive coefficient of Infant Mortality Rate, an effect associated with the physiological frailty of boy infants rather than with any difference in child care practices or local conditions. It might be reasonable to allow the coefficient of this variable to differ across the two equations before testing for structural stability. Once again the unrestricted coefficient vector relates to separate regressions, but now the restriction on IMR is dropped in the semi-pooled regression. The F-ratio that results from estimating this amended model is reported in the first panel of Table 14.6. The hypothesis of structural stability cannot now be rejected, suggesting that the difference in the coefficients associated with IMR accounts for almost all of the difference in the coefficient vectors.

Overall then, the analysis confirms the view that different mortality in the past was not the outcome of discrimination in childhood associated with parental preferences, or anticipation of economic returns, or social norms. Instead it was a developing phenomenon intertwined with increasingly marked divisions of labor and sharply distinguished tasks and responsibilities.

V. Conclusions

Following Amartya Sen's pioneering work on "missing women", development economists have publicized the excess deaths of female children, interpreted as the functional equivalent of female-intensive infanticide, in the context of strong son preference deriving mainly from the support which sons afford in old age. The historical story is different. The analysis offered in this chapter confirms earlier evidence that shifts attention away from childhood, the phase of life of most concern to development economists. Not only did little girls not die at higher rates than their male peers in the aggregate, their death rates varied in the same ways and were affected by the same variables as little boys, and the same is true of girls aged 5 to 9. But teenagers and young women did suffer excess female mortality. To read this phenomenon as "deferred infanticide" would cast nineteenth-century parents as inefficient as well as inhumane. Moreover, as female death rates diverged

from those of their male peers they appeared to be differently affected by local economic and environmental factors. It is true that maternal mortality and an apparent susceptibility to tuberculosis were factors that elevated death rates in these age groups, but they cannot account quantitatively for the excess deaths or provide an explanation. Instead, our interpretation has emphasized the different responsibilities and roles that fell on teenagers and young adults as they grew older. Once girls reached working age, employment became relevant not merely because of the enhanced status afforded by working but also because of the physical wear and tear which some jobs imposed. For older girls the marriage market was implicated in both the determination of worth and perhaps the mechanisms whereby mortality was elevated. Youthful marriage suggests that girls had few alternatives, and it simultaneously exposed them to early pregnancy with its high risks.

For adult women, the nexus of social and economic forces which influenced life chances was different yet again, with motherhood and its culturally demanded behavior paramount. The positive and significant role of fertility in explaining adult women's mortality underlines the direct hazards of childbearing, the longer-term strains of reproductive work, and the importance of social norms of maternal sacrifice in the context of many small children and scarce resources.

Another important lesson for modern feminists is that the labor market offered little leverage on life chances. Working-class districts, where women were more likely to have contributed economically, were less, not more, favorable for women than more middle-class districts, where women were more likely to have retreated into domesticity and dependence. This may reflect the difficulties that married women bore in combining waged work with family life. In nineteenth-century conditions the double shift was an intolerable burden. The long hours, lack of rest, poor diet, and simple overwork associated with combining employment and domestic and reproductive activities took their toll.

Thus the origins of differences in mortality in the past shifts the research agenda away from fixation on child care and parenting to the terms and conditions of employment, the nature of the fertility regime and, perhaps most important of all, the sacrifices demanded of nineteenth-century mothers. Here the historical evidence chimes well with a theme that is strongly evident in Amartya Sen's work, for Sen has emphasized the importance of values and "positional" observations in sustaining gender inequity (Sen 2006). An attitudinal fog of the kind Sen seeks to disperse undoubtedly confounded the perceptions of nineteenth-century mothers, causing them to participate in the practices that undermined their health and eroded their well-being. Perhaps it remains to cloud the views of many twenty-first-century historians, who want to blame parents or preferences and not power structures for the ugly remnants of excess female mortality.

References

Agarwal, B. (1997). " 'Bargaining' and Gender Relations: Within and Beyond the Household", *Feminist Economics*, 3: 1–51.

Anderson, M. (1990), "The Social Implications of Demographic Change", in F. M. L. Thompson (ed.), *The Cambridge Social History of Britain 1750–1950*, vol. 2 (Cambridge: Cambridge University Press), 1–70.

Bardhan, P. (1974), "On Life and Death Questions", *Economic and Political Weekly*, 9: 1283–304.

Bourke, J. (1994), "Housewifery in Working-Class England 1860–1914", *Past and Present*, 142: 167–97.

Bristow, B. R. (1985), "Population and Housing in Nineteenth-Century Lancashire: A Framework for Investigation", *Local Population Studies*, 34: 12–26.

Bryder, L. (1988), *Below the Magic Mountain: A Social History of Tuberculosis in the Twentieth Century* (Oxford: Clarendon Press).

Burnette, J. (1997), "An Investigation of the Female–Male Wage Gap During the Industrial Revolution in Britain", *Economic History Review*, 50: 41–67.

Chen, L. C., Huq, E., and D'Souza, S. (1981), "Sex Bias in the Family Allocation of Food and Health Care in Rural Bangladesh", *Population and Development Review*, 7: 55–70.

Daly, M., and Wilson, M. (1988), *Homicide* (London: Transaction Publishers).

Das Gupta, M. (1987), "Selective Discrimination Against Females in Rural Punjab, India", *Population and Development Review*, 13: 77–100.

Devos, I. (1996), "La regionalisation de la surmortalité feminine des jeunes filles en Belgique entre 1890 et 1910", *Annales de Démographie Historique*, 33: 375–407.

—— (1998), "Excess Female Mortality at the End of the Nineteenth Century in Northern France and Belgium", paper presented at the 23rd Annual SSHA Conference, 19–22 November, Chicago.

D'Souza, S., and Chen, L. C. (1980), "Sex Differences in Mortality in Rural Bangladesh", *Population and Development Review*, 6: 257–70.

Eggerickx, T., and Tabutin, D. (1994), "La surmortalité des filles vers 1890 en Belgique: une approche régionale", *Population*, 49: 657–84.

Farr, W. (1885), *Vital Statistics: A Memorial Volume of Selections from Reports and Writings of William Farr* (London: Office of the Sanitary Institute).

Ginsberg, C. A., and Swedlund, A. C. (1986), "Sex-Specific Mortality and Economic Opportunities: Massachusetts, 1860–1899", *Continuity and Change*, 1: 415–45.

Harris, B. (forthcoming), "Gender and Wellbeing", *Research in Economic History*.

Humphries, J. (1977), "Class Struggle and the Persistence of the Working-Class Family", *Cambridge Journal of Economics*, 1: 241–58.

—— (1991), " 'Bread and a Pennyworth of Treacle': Excess Female Mortality in England in the 1840s", *Cambridge Journal of Economics*, 15: 451–73.

—— and McNay, K. (1998), "Gender, Death and Welfare During the British Industrial Revolution", unpubl. MS, All Souls College, Oxford.

Johansson, S. R. (1977), "Sex and Death in Victorian England: An Examination of Age- and Sex-Specific Death Rates 1840–1910", in M. Vicinus (ed.), *A Widening Sphere* (Bloomington, Ind.: Indiana University Press), 163–81.

Johansson, S. R. (1991), "Welfare, Mortality and Gender: Continuity and Change in Theories about Male/Female Mortality Differences Over Three Centuries", *Continuity and Change*, 6: 135–77.

Klasen, S. (1998), "Marriage, Bargaining and Intrahousehold Resource Allocation: Excess Female Mortality Among Adults During Early German Development, 1740–1860", *Journal of Economic History*, 58: 432–67.

—— and Wink, C. (2005), " 'Missing Women': Revisiting the Debate", in B. Agarwal, J. Humphries and I. Robeyns (eds), *Amartya Sen's Work and Ideas: A Gender Perspective* (Abingdon, Oxon.: Routledge), 265–302.

McNay, K., Humphries, J., and Klasen, S. (1998), *Death and Gender in Victorian England and Wales: Comparisons with Contemporary Developing Countries*, Working Paper 9801 (Cambridge: Department of Applied Economics, Cambridge University).

—— —— —— (2005), "Excess Female Mortality in Nineteenth-Century England and Wales: A Regional Analysis", *Social Science History*, 29: 649–81.

Mitchell, B., and Deane, P. (1962), *Abstract of British Historical Statistics* (Cambridge: Cambridge University Press).

Oddy, D. (1976), "A Nutritional Analysis of Historical Evidence: The Working-Class Diet 1880–1914", in D. Oddy and D. Miller (eds), *The Making of the Modern British Diet* (London: Croom Helm).

Oren, L. (1974), "The Welfare of Women in Labouring Families in England, 1860–950", in L. Banner and M. Hartman (eds), *Clio's Consciousness Raised: New Perspectives on the History of Women* (New York: Harper Colophon).

Rosenzweig, M. P., and Schultz, P. T. (1982), "Market Opportunities, Genetic Endowments, and Intra-Family Resource Distribution", *American Economic Review*, 72: 803–15.

Ross, E. (1982), " 'Fierce Questions and Taunts': Married Life in Working-Class London, 1870–1914", *Feminist Studies*, 8: 575–98.

Sen, A. (1990), "Gender and Co-Operative Conflict", in I. Tinker (ed.), *Persistent Inequalities: Women and World Development* (Oxford: Oxford University Press), 123–49.

—— (2006), "Continuing the Conversation", in B. Agarwal, J. Humphries and I. Robeyns (eds), *Capabilities, Freedoms and Equality: Amartya Sen's Work from a Gender Perspective* (Delhi: Oxford University Press), 347–61.

Tabutin, D. (1978), "La surmortalité féminine en Europe avant 1940", *Population*, 33: 121–48.

—— and Willems, M. (1998), "Differential Mortality by Sex from Birth to Adolescence: The Historical Experience of the West (1750–1930)", in United Nations (ed.), *Too Young to Die: Genes or Gender*? (New York: United Nations), 17–52.

United Nations (1998), "The Extent and Causes of Female Disadvantage in Mortality: An Overview", in United Nations (ed.), *Too Young to Die: Genes or Gender*? (New York: United Nations), 1–15.

Waldron, I. (1983), "The Role of Genetic and Biological Factors in Sex Differences in Mortality", in A. D. Lopez and L. T. Ruzicka (eds), *Sex Differentials in Mortality* (Canberra: Australian National University Press), 141–64.

Wall, R. (1981), "Inferring Differential Neglect of Females from Mortality data", *Annales de Demographie Historique*, 18: 119–40.

Willems, M. (1993), "Realité et diversité de la surmortalité féminine dans le monde et dans l'histoire", unpubl. Master's diss., Institut de Démographie, Université de Louvain-la-Neuve.

Woods, R., and Hinde, P. R. A. (1987), "Mortality in Victorian England: Models and Patterns", *Journal of Interdisciplinary History*, 18: 27–54.

—— and Shelton, N. (1997), *An Atlas of Victorian Mortality* (Liverpool: Liverpool University Press).

Wrigley, A. E., Davies, R. S., Oeppen, J. E., and Schofield, R. S. (1997), *English Population History from Family Reconstitution 1580–1837* (Cambridge: Cambridge University Press).

CHAPTER 15

MISSING WOMEN

SOME RECENT CONTROVERSIES ON LEVELS AND TRENDS IN GENDER BIAS IN MORTALITY

STEPHAN KLASEN

I. Introduction

One of Amartya Sen's important contributions in the field of development economics has been his work on gender bias in mortality in parts of the developing world. Although imbalances in aggregate sex ratios favoring males had been known for some time (see e.g. Visaria 1961; Bardhan 1974), he added to this literature by providing new empirical evidence on gender discrimination in India (Sen and Sengupta 1983), by suggesting a modeling framework that has powerfully influenced the literature on intra-household resource allocation (Sen 1990a), by coining the term "missing women", referring to the cumulative impact of past and present gender bias in mortality on the current sex structure of the population, and by

I would like to thank Jason Abrevaya, Pranab Bardhan, Monica Das Gupta, Brad Delong, Ravi Kanbur, Ling-Jen Lin, Emily Oster and Amartya Sen for helpful comments and discussion.

providing a first estimate of the number of missing women in the countries affected by gender bias in mortality (Sen 1990b). In well-known articles in 1989 and 1990, he suggested that more than 100 million women are "missing" in South Asia, East Asia, the Middle East and North Africa, and that they have fallen victim to gender bias in mortality (Sen 1989, 1990b). He arrived at this figure by comparing the sex ratios of these countries and regions with those prevailing in Sub-Saharan Africa and assuming that the difference was due to gender bias in mortality. A literature subsequently developed that has refined these estimates of missing women, extended them to more countries, and updated them using more recent census information (see e.g. Coale 1991; Klasen 1994; Klasen and Wink 2002, 2003; OECD 2006). A much larger literature has examined gender bias in mortality in South and East Asia in much greater detail to understand its demographic correlates as well as its socioeconomic causes (see e.g. Drèze and Sen 2001 for a summary of many of these contributions). Due to its obvious importance from a policy point of view, this literature has continued to grow.

One issue that has sparked a particular amount of debate has been the effect of fertility decline on gender bias in mortality. In particular, several scholars have suggested that fertility decline in countries with son preference leads to intensified gender bias in mortality (e.g. Das Gupta and Mari Bhat 1997; Basu 1999, 2000; Rajan *et al.* 2000). Others have suggested, however, that this link is empirically weak, at least in India, or appears actually to go in the opposite direction (e.g. Drèze and Murthi 2001; Mari Bhat and Zavier 2003). The first purpose of this chapter is to reflect on this debate using some recent data from countries affected by gender bias in mortality. In particular, I will suggest that there is no evidence of a solid relationship between fertility decline and intensification of gender bias; in certain circumstances, particularly those where fertility decline is partly a result of coercive family planning policies (as in China) or where son preference has remained strong but desired fertility is already very low (as in South Korea), fertility decline can lead to an intensification of gender bias, at least for some length of time; in most other circumstances, however, the evidence for such a link is weak, either across or within nations. In fact, often the opposite appears to be the case: fertility decline goes hand in hand with declines in gender bias in mortality.

The last two years have also seen a return of debates on the magnitude of the number of "missing women". In particular, a paper by Emily Oster (2005) has claimed that existing estimates of "missing women" vastly overstate the phenomenon, as they neglect a link between the Hepatitis B virus (HVB) and the sex ratio at birth. She argues that there is a positive link between the proportion of parents who are carriers of HVB and the sex ratio at birth. Since many of the countries with gender bias in mortality also have (or at least have had) a fairly substantial prevalence of HVB carriers, about 50% of the "missing women" alleged to be victims of gender bias in mortality are missing, more innocuously, as a result of the HVB carrier status of their parents. This claim has also sparked a controversy,

with most of the evidence presented by Oster having come under critical scrutiny and other conflicting evidence has been put forward (see Abrevaya 2005; Das Gupta 2005, 2006; Lin and Luoh 2008; Ebenstein 2007). In fact, Oster herself has very recently produced evidence that goes against her original claims and she has therefore retracted the claim that HVB prevalence was responsible for bias sex ratios, at least in China (Chen and Oster 2008). The second aim of the paper is to comment on the original claims made by Oster as well as the literature that has developed since. I will argue that while the claim was rather intriguing and the empirical evidence presented by Oster substantial, closer inspection of that evidence generates serious questions and problems with *all* pieces of the evidence presented in support of this claim. Moreover, the evidence against a link between HVB carrier status and the sex ratio at birth, as well as the evidence on parental strategies of son preference, is so substantial that the claim that much of gender bias in mortality is mostly due to this biological linkage appears to be largely baseless. It appears instead that any possible link between HVB carrier status and the sex ratio at birth has at most a tiny impact on the number of "missing women". As a result, it remains the case that gender bias in mortality (including pre-birth strategies of sex selection) accounts for an overwhelming portion of the sex ratio imbalances observed in South and East Asia.

II. Fertility Decline and Gender Bias in Mortality

Most of the countries affected by gender bias in mortality have undergone a significant fertility decline in the past 10 to 20 years. In fact, as shown in Table 15.1, which presents total fertility rates (TFRs) matched to the two most recent censuses, which have been used for the missing women calculations in Klasen and Wink (2003), TFRs have fallen dramatically in many nations, including Bangladesh, India, Pakistan and most Middle Eastern countries. In China and Korea, much of the fertility decline occurred prior to the 1990s, and is now below replacement levels. Only in Afghanistan has fertility persisted at extremely high levels.

There are a number of reasons why one might expect such a fertility decline to lead to an intensification of gender bias in mortality. Given an unchanging desire of many parents in the regions affected by gender bias in mortality to have at least one surviving son, a lower TFR could mean that parents will be particularly concerned that the one surviving son be male. This is called the intensification effect by Das Gupta and Mari Bhat (1999). In addition, Basu (1999) argues that son preference

Table 15.1. Fertility decline and changes in gender bias in mortality

	Most recent data	Females "missing" (%)	TFR	Previous data	Females "missing" (%)	TFR	Decline in TFR	Change in percentage missing
Afghanistan	2000	9.30	6.84	1979	9.70	6.9	0.06	0.40
Algeria	1998	1.24	3.05	1987	2.69	5.72	2.67	1.45
Bangladesh	2001	4.23	3.13	1981	8.91	6.0	2.87	4.68
China	2000	6.69	1.89	1990	6.31	2.1	0.21	−0.38
Egypt	1996	4.46	3.55	1986	5.14	4.3	0.75	0.68
India	2001	7.89	3.07	1991	9.44	3.81	0.74	1.55
Iran	1996	3.68	3.0	1986	4.54	5.9	2.9	0.86
Korea	1995	0.73	1.75	1985	−0.14	2.04	0.29	−0.87
Nepal	2001	0.50	4.27	1981	7.71	6.1	1.83	7.22
Pakistan	1998	7.76	4.77	1981	10.78	7.0	2.23	3.01
Sri Lanka	1991	−0.02	2.5	1981	3.44	3.25	0.75	3.47
Syria	1994	3.06	4.2	1981	4.98	7.4	3.2	1.92
Tunisia	1994	2.13	2.9	1984	4.45	4.32	1.42	2.33
Turkey	1990	2.40	3.0	1985	3.15	3.79	0.79	0.75

Sources: Klasen and Wink (2003); World Bank (2007).

might even increase as a result of fertility decline, further intensifying gender bias.[1] As a result, the whole range of pre- and post-birth strategies could be applied to ensure that the far fewer children include at least one male. Among the strategies parents would have at their disposal are selective stopping rules, i.e. ending child-bearing as soon as the first son is born, using sex-selective abortions to increase the likelihood of sons, and favoring boys in the allocation of survival-related resources such as health and nutrition (see Klasen 2003). Of course, all three strategies could be applied at the same time. It is likely that the effects on sex ratios would be particularly severe in countries with substantial son preference where the fertility decline is a result of coercive family planning policies (such as China), or where fertility levels are so low that a large share of parents would not have a surviving son unless some of these strategies to manipulate the sex of the offspring were applied (such as Korea).

While these arguments are all plausible, they treat fertility decline as an exogenous variable affecting gender bias in mortality, assuming unchanging (or even increasing) son preference. Given the literature on the determinants of fertility levels, however, it is much more reasonable to treat fertility decline as an endogenous variable affected by a whole set of developments in a country, including rising prosperity and rising levels of female education, which in turn also have an impact

[1] See Basu (1999) for a discussion of the possible reasons for increased son preference.

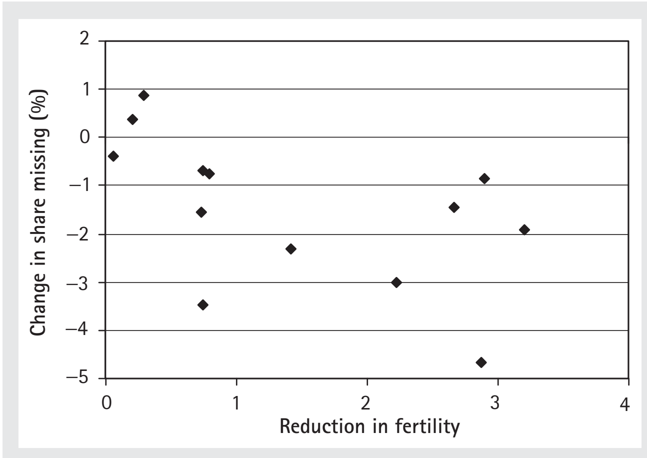


Fig. 15.1. Fertility decline and changes in the share of missing females

Source: Data from Table 15.1.

on son preference and thus on gender bias in mortality.[2] Thus one might plausibly argue that fertility decline and gender bias in mortality are jointly determined by these developments. In that case, it could well be that endogenous fertility declines are associated with *reductions* in gender bias in mortality, in particular if the causal factors driving these developments are rising incomes, rising female education, and rising female employment levels, all of which are likely to be associated with reduced gender bias in mortality.[3]

As a result, the relation between fertility decline and gender bias in mortality is a complex empirical question. In the following I want to present briefly some empirical evidence on this relation, both across countries and within selected countries, including China, Korea and India, where these issues have been particularly prominent.

Table 15.1 and Figure 15.1, which plots the two rightmost columns of Table 15.1, speak rather clearly on the relationship between fertility decline and gender bias in mortality in the countries with the most serious problem of "missing women". Table 15.1 reports on the share of "missing females" in the last two censuses as an indicator of (pre- and post-birth) gender bias in mortality in the "missing women" countries, as calculated by Klasen and Wink (2003), and relates this to absolute fertility decline over the same period. If the fears about an intensification of gender bias as a result of fertility decline were correct, we would expect a positive relationship between fertility decline and change in gender bias in mortality, i.e. the larger the fertility decline, the higher the increase in gender bias in mortality. Figure 15.1 quite clearly

[2] See, for example, Murthi *et al.* (1995), Drèze and Murthi (2001) and Bhattacharya (2006) for a discussion in the Indian context.

[3] For a discussion of these links from both theoretical and empirical standpoints, see Klasen and Wink (2002).

shows nothing of the sort. In fact, the correlation goes the other way and is actually highly significant. In countries where fertility decline has been largest, the share of missing women has tended to fall the most.[4]

China and (South) Korea are included in Table 15.1 and Figure 15.1 and therefore one might plausibly object that in these cases most of the fertility decline had taken place before 1990 (1985 in the case of Korea), so it could have been the previous, stronger fertility decline that is responsible for an intensification of gender bias. In the case of China, there clearly is some force to this argument. As demonstrated very clearly in Banister (1987), Banister and Coale (1994), Lai (2005), Attane (2007), and Ebenstein (2007), the fertility decline since the late 1970s has been associated with a sharp rise in both pre- and post-birth discrimination against female children, leading to a sharp rise in the share of missing females in China since the early 1980s. In the 1990s, fertility decline slowed considerably and gender bias in mortality has not deteriorated much further from the already poor state of affairs it had reached in 1990. But it should be pointed out that the fertility decline in China is of a special sort, as it has been heavily affected by the one-child policy instituted in 1977, where parents have to contend with severe sanctions if they have more than one child. In contrast to other countries, fertility decline in the context of the coercive one-child policy can be treated as largely "exogenous".[5] In an environment of significant remaining son preference, it is therefore not surprising that some of the intensification effects predicted for India have been found in China, at least in the initial phases of the one-child policy.

It is unclear, however, whether this large increase in gender bias in mortality in China since the introduction of the one-child policy is entirely due to the fertility decline the latter generated. While there is no doubt that the relative deterioration in female survival is largely due to the effects of the one-child policy, the sharp rise in sex-selective abortions in China since the mid-1980s (increasing the male–female ratio at birth from about 1.06 in the early 1980s to 1.20 in 2000: see Klasen and Wink 2003 and Attane 2007) might be only partly related to the one-child policy. Instead, it is possible to view this rising incidence of sex-selective abortions as an exogenous technological change (brought about by the increasing availability of pre-natal sex determination through ultrasound screening) that has enabled parents to achieve their desired sex composition of their offspring more easily than had been possible.[6]

[4] In Figure 15.1, Nepal is omitted from the scatterplot. The change in the share of missing females in Nepal between 1981 and 2001 was strongly affected by sex-specific out- and in-migration patterns, so the calculation of "missing females" for this country has to be treated with some caution. Even if it were included, however, there would remain a clear negative relationship between fertility decline and the change in the share of missing women (although the correlation would be somewhat weaker).

[5] It is, of course, debatable to what extent the fertility decline in China has actually been driven by the one-child policy. Clearly, significant fertility decline was well under way before the one-child policy was instituted. But it is also plausible that the pace and extent of fertility decline were significantly influenced by that policy. For discussion, see Sen (1998) and Drèze and Sen (2001).

[6] See Klasen and Wink (2002, 2003), Mari Bhat and Zavier (2003) and Chung and Das Gupta (2007) for a more detailed discussion and for evidence supporting this claim.

South Korea is an interesting test for this argument. Here fertility decline took place without recourse to draconian family planning policies; in addition, pre-birth sex determination has been available since the 1980s. As shown by Chung and Das Gupta (2007), the male–female ratio at birth rose sharply from the early 1980s to the early 1990s (from about 1.07 to 1.15), coinciding with the increasing availability of pre-birth sex determination. This supports the claim that a change in technology alone can severely affect gender bias in mortality in an environment of strong son preference and falling or low fertility.

The Korean story does not end there. In the midst of rising education, modernization, economic development and continued fertility decline, the sex ratio at birth has been falling since the early 1990s and at 2003 was close to its levels in the early 1980s. As shown by Chung and Das Gupta (2007), the main driver of this change is falling son preference associated with a secular change in attitudes brought about by economic and social development. Thus after a rise in the male–female ratio at birth brought about by technological change, Korea now fits the pattern described above, where economic and social development appears to reduce both fertility and son preference. Chung and Das Gupta expect a similar trend to happen in due time in China (and India).

Thus the Chinese experience suggests that fertility decline resulting from coercive family planning can powerfully intensify gender bias in mortality. At the same time, not all of the intensification of gender bias in mortality is due to this, as technological changes in China and Korea have helped to increase the ability of parents to determine the sex composition of their offspring.

Let me now turn to India. One way to examine the relationship between fertility decline and gender bias in mortality for India is to examine this relationship at the state level. Table 15.2 presents some simple regressions of the statewise sex ratio (males/females) in India's largest states between 1961 and 2001. I show pooled-OLS, fixed-effects and random-effects results.[7] The pooled-OLS results suggest that high fertility is associated with a higher sex ratio, i.e. a higher share of males, at the state level, contrary to the hypothesis examined here. It is unlikely, however, that high fertility is "causing" the high sex ratio; rather the fertility rate is an endogenous variable, with some underlying conditions affecting both fertility as well as gender bias at the state level. This suspicion is confirmed by the random effects and, in particular, the fixed-effects specification. With the latter, one would hope that all state-specific factors affecting the sex ratio at birth are captured by the fixed effects and only the impact of the change of fertility on the sex ratio is being examined. The results show that there is no longer a positive significant relationship, but neither is there a negative relationship. So the results for India also demonstrate that there is no evidence of fertility decline leading to an intensification of gender bias in mortality.

[7] Specification tests suggest that the random-effects results are to be preferred.

Table 15.2. Determinants of the sex ratio (males/females) of India's largest states, 1961–2001

	OLS	Random effects	Fixed effects
Total fertility rate	1.91 (0.86)	0.29 (0.62)	0.09 (0.16)
South	−6.15 (−3.09)***	−8.53 (−3.21)***	
East	−3.68 (−2.42)***	−3.88 (−1.14)	
West	−1.32 (−0.86)	−2.72 (−0.88)	
1971	−4.17 (−1.86)**	−0.07 (−0.05)	0.39 (0.28)
1981	−2.77 (−1.62)*	−0.49 (−0.56)	−0.21 (−0.24)
1991	−0.32 (0.75)	0.39 (0.75)	0.49 (1.07)
Adjusted R^2	0.52	0.54	0.13
n	58	58	58

Notes: *t*-ratios based on robust standard errors are reported in parentheses. * 90%, ** 95% and *** 99% significance level (one-tail test). For details of the states included, see Klasen and Wink (2003).
Source: Own analysis based on data from Registrar General (2003).

Other research also supports the contention that fertility decline has not led to an intensification of gender bias in India. In particular, Mari Bhat and Zavier (2003) examine changes in son preference and fertility decline in India using the Indian National Family Health Surveys. They find that son preference has declined in the course of fertility decline in every state but one. This is further evidence that one cannot treat fertility decline as an exogenous factor affecting gender bias in mortality, holding son preference fixed. It seems much more plausible that the factors driving fertility decline, including rising female education, employment and income, are precisely the factors that also reduce son preference.

In the Indian case, however, it is also likely that the increasing availability of sex-determination technologies is having a one-off impact on the level of sex-selective abortions and thus pre-birth gender bias in mortality. Thus the worrying increase of the sex ratio for 0- to 6-year-olds in the 2001 census is likely to be related mostly to this one-off technological shift, which has increased the ability of parents to affect the sex of their offspring, rather than a result of fertility decline. This can also be seen by examining the year dummies in the regressions in Table 15.2. Compared to 2001, the sex ratio in prior decades was always lower, so there appears to be a secular upward trend, particularly since the 1981 census. This would be perfectly consistent with the impact of such a technological shift.[8] If this argument is correct, the prediction would be that the proportion of males in the younger age groups should slowly drift downwards together with declining son preference, as it recently has in South Korea.

[8] See also Klasen and Wink (2003) and Mari Bhat and Zavier (2003) for a similar argument.

III. Hepatitis B and the Number of Missing Women

Let me turn now to the second debate about levels and trends in gender bias in mortality. As noted earlier, Emily Oster (2005) suggested that a considerable share of the "missing women", particularly in China, are not due to gender bias in mortality, but to the prevalence of hepatitis B virus (HVB) carriers among the adult population, which, she argued, has been found to increase the proportion of males at birth. Females have not died through pre- or post-birth sex-selection strategies, but were never born due to this epidemiological link; thus many "missing women" are due to "biology" rather than "discrimination", a finding that has also been widely discussed in the popular press (see e.g. Barro 2005). As shown in Table 15.3, reproduction of a table from her paper, she believed that about half of the estimated 60 million missing women in Coale's (1991) calculation (based on census data available at the time) can be explained this way. The impact on the estimates of missing women is largest in China (a reduction of −77%), where the problem of gender bias in mortality largely vanishes, followed by Egypt (−54%) and West Asia (−30%), and the impact is lowest in South Asia (−18% in India, −20% in Bangladesh, and no impact in Pakistan).

As shown in Table 15.3, Oster also compared her estimates of missing women with those of Sen (1992), whereupon the share of missing women supposedly explained by this biological link rises to 70%. This particular comparison makes little sense, however. In contrast to Coale, Sen's estimates are not based on an explicit assumption about the sex ratio at birth. Instead, he generated his figures by simply

Table 15.3. Missing women reported by Sen and Coale and "explained" by Oster

	Coale (1991)			"Coale updated"	Sen (1992)		
	Missing	Adjusted missing (m)	Share "explained" (%)	Missing	Missing	Adjusted missing (m)	Share "explained" (%)
China	30.4	7.2	76.4	32.3	50.0	7.2	85.6
India	22.7	18.6	18.2	24.6	41.6	18.6	55.3
Pakistan	3.1	3.1	0.3	3.4	5.2	3.1	40.6
Bangladesh	1.6	1.3	19.6	1.0	3.8	1.3	65.7
Nepal	0.2	0.2	0.0	−0.3	0.6	0.2	60.8
Western Asia	1.6	1.1	29.5	1.8	4.7	1.1	75.8
Egypt	0.6	0.3	53.7	0.7	1.7	0.3	82.9
Total	60.3	31.8	47.2	63.6	107.5	31.8	70.4

Source: Oster (2006). "Coale updated" is taken from Klasen and Wink (2003), using his method and applying it to census data of the late 1990s and early 2000s.

comparing the sex ratios in the *whole population* (not the sex ratios at birth) in the "missing women" countries with those in Sub-Saharan Africa. In fact, if there is a link between HVB carrier status and sex ratio at birth, there is no need to adjust population sex ratios from Sub-Saharan Africa to account for this link, as its rate of HVB carriers is similar to that of China.[9] Nonetheless, if Oster's analysis of Coale's figures is sound, she has just about defined away the problem in China and sharply reduced it in Egypt, West Asia and most of South Asia, so it is important to assess the empirical evidence for her analysis.

Before doing this, it is useful, however, to comment briefly on one important implication of her results if they proved correct. Since the 1980s, most countries affected by gender bias in mortality, including China, Taiwan and parts of South Asia, have successfully undertaken massive immunization campaigns to reduce the prevalence of HVB, thereby reducing the carrier rate drastically among young couples, which should in turn increasingly affect the sex ratio at birth (and, more slowly, the population sex ratio). If there is a link between HVB carrier status and the sex ratio at birth, this should serve to lower the sex ratio at birth (and of the population as a whole) over time; these effects should have become strongly visible in the 1990s and have affected the sex ratios of the censuses around 2000. Consequently, in the most recent censuses, the gap between the actual sex ratio at birth and the HVB-free sex ratio at birth disappears or at least becomes much smaller; consequently, the number of missing women reported in these censuses increases. One way to gauge the quantitative impact is to examine the estimate of the number of missing women, using Coale's methods and census data from the late 1990s and early 2000s (which I shall call "Coale updated", calculated in Klasen and Wink 2003). Using these comparable methods but not accounting for the alleged link between HVB and the sex ratio at birth implies that the number of missing women went up only slightly by about 3 million and fell as a share of the female population (from 5.3% around 1990 to 4.7% around 2000). If, however, the figures in Coale (1991) were inflated due to the neglect of this link, as suggested by Oster, but are no longer inflated in 2000 due to the success of immunization campaigns, the number and share of missing females increased dramatically in the 1990s, from 32 million to 64 million (and from about 2.8% to 4.7% of the female population). This is a rather extreme assumption to illustrate the argument. Even if the immunization campaigns had been totally successful by the early 1990s in eradicating HVB carrier status among mothers, the effect of HVB on the sex ratio at birth in older cohorts would still be present. But the main point, that the improvements found by Klasen and Wink (2002, 2003) would vanish and turn into a deterioration, remains. Thus

[9] There are other reasons why the use of the sex ratio in Sub-Saharan Africa is unlikely to produce reliable estimates of the number of missing women. These are discussed in detail in Klasen (1994) and Klasen and Wink (2002, 2003). Nevertheless, the unusually low male–female ratios in Sub-Saharan Africa, despite very high HVB carrier prevalence, clearly works against the hypothesis proposed by Oster. See the discussion below.

while Oster's findings might be comforting in that the *levels* of gender bias in mortality were smaller in the early 1990s than commonly thought, her results imply that the problem has since got a lot worse than previously believed. This rather worrying implication was not mentioned in her paper.

There are two principal ways to assess the merits of her evidence. One is to examine the different pieces of evidence to see whether they make a convincing case when assembled. The other is to look for evidence on this linkage that has not been put forward by Oster. In the following, I will do both briefly.

Oster begins by presenting time series evidence on the sex ratio at birth and in young age groups in countries affected by gender bias in mortality, with particular emphasis on China. The data on sex ratios in young age groups are difficult to assess given the incentive of parents to under-report children (particularly female children) after the introduction of the one-child policy in 1977.[10] The data on the sex ratio at birth are based on a careful historical reconstruction from Banister and Coale (1994) using data from fertility surveys.[11] While on average they indeed show elevated (male–female) sex ratios at birth on average and for most time periods since 1930 (compared to the international norm of about 1.06), they oscillate considerably and there are periods in the 1960s where the sex ratio at birth is not much different from the norm in Western countries. As there is no evidence for similar swings in HVB prevalence, this already casts doubt on the presumed link. Also, as these data are based on retrospective surveys, one has to wonder whether in a country with strong son preference there is a sex-specific recall bias. Consistent with this view is the evidence that the sex ratios reported are larger the older the birth cohort.

More seriously, more disaggregated data show particular groups without elevated sex ratios at all. For example, data from the 1982 census on sex ratios at birth (reported in Lai 2005) show that in 11 provinces (out of 29 for which data are reported) the sex ratio at birth is close to 1.06 or below and thus very similar to the international norm. Similarly, the same paper by Banister and Coale (1994) shows that among first-born children, the sex ratio was within the international norm of about 1.06 for the birth cohorts between 1960 and 1980. Das Gupta (2005) also reports that in 1982 and 1989 the sex ratio at birth in China was equal to the international norm for the first-born child. Both Banister and Coale and Das Gupta find a sharply increased sex ratio at birth only for later-born children, and the sex ratio among these later-born children has increased sharply since the early 1980s. Lastly, Ebenstein (2007), in the most comprehensive assessment of the issue, also presents evidence of the sex ratio at birth in China, India and Taiwan from 1980 to 2000. He also finds that sex ratios of first births are perfectly normal and only rise for higher parities, particularly since 1990, consistent with increasing recourse to sex-selective abortions. All this evidence casts doubt on the claim that sex ratios

[10] See Johansson and Nygren (1991) for a careful empirical assessment of the hiding of daughters in China.

[11] These data are not affected by the incentive to hide daughters associated with the one-child policy.

at birth are in general higher in China due to HVB carrier status. The only way this evidence would be consistent with a link to HVB carrier status is if carrier status only affected higher-order parities, but not first births. It is hard to think of a way this could be possible, given that all higher-order births by definition imply lower-parity prior births and that HVB carrier status in high-prevalence countries is usually determined in childhood well before the beginning of childbearing. Also, the microevidence discussed below suggests precisely the opposite relationship between HVB, parity, and the sex ratio at birth.

A second piece of evidence on elevated sex ratios at birth related to HVB carrier status among Chinese populations comes from Chinese immigrants in the USA. According to census data, the average sex ratio among Chinese immigrants is 1.082 when birth registry data are used and 1.105 when smaller census samples are used, which Oster claims to be related to higher HVB prevalence among Chinese immigrants, whose carrier status is affected by conditions in China prior to immigration. Given the source, these data are likely to be much more reliable. But closer inspection also reveals two problems. First, as shown in Table 2 in Oster's paper, between 1940 and 1970 the sex ratio of children born to Chinese immigrants was perfectly normal at 1.047. Only much further back and since 1980 has the sex ratio at birth been abnormally high.

Related to this is a second problem identified in a recent paper by Abrevaya (2005). Using the same data as Oster plus a richer subsample of all births from California, he first shows that among Chinese, Indian and South Korean immigrants, there is significant gender bias in fertility strategies. The likelihood of these immigrant groups to have a second or third child is significantly higher if the first child (or the first two) has been a girl than if it has been a boy (or two boys). Possibly as a result of fertility decline, this link has become somewhat stronger since 1980, thereby in fact connecting to the discussion above about the impact of fertility decline on gender bias. While this is an interesting finding in itself, such a selective stopping rule does not affect the sex ratio at birth in the absence of sex-selective abortions (see Klasen 2003). But Abrevaya's paper also reports that the actual sex ratio at birth is affected by family size and the sex composition of the family. Table 15.4 reports some descriptive statistics from that paper on the sex of

Table 15.4. Sex ratio of second child by sex of first child, USA, 1980–1995

	Girl	Boy
Chinese	1.110	1.033
Indian	1.123	1.016
Korean	1.070	1.033

Source: Abrevaya (2005).

the second child of Chinese, Indian and Korean immigrants if the first child was a girl or a boy. In each case, the sex ratio is significantly (in a statistical and economic sense) higher when the first child was a girl than when it was a boy. This relationship also holds up in regressions that control for other factors affecting the sex ratio at birth. This is rather strong evidence that the higher sex ratios observed in these Asian immigrant groups are also due to sex-selective abortions of parents trying to ensure that at least one of their children is male. Abrevaya (2005) also presents a range of circumstantial evidence supporting this contention. Conversely, these data, particularly the perfectly normal sex ratios of births when the first child was male, provide strong evidence against the hypothesis that HVB carrier status causes the abnormally high sex ratios among immigrants in the US. If such a link existed, one would expect a generally higher sex ratio at birth and not a link between the sex ratio at birth and parity.

Oster then turns to individual and time series evidence for further support. The individual-level evidence reports on the results of several small microstudies that particularly examine this link, while the time series evidence reports on the impact of vaccination campaigns on populations heavily affected by HVB carrier status. The individual-level evidence finds a strong and significant relationship between HVB carrier status and the sex ratio at birth, while the time series evidence finds that for Alaskan Natives as well as Taiwanese, vaccination campaigns are associated with reduced sex ratios at birth.

Let me raise a few empirical issues regarding the microstudies. The sample sizes from the microstudies are very small and the results thus quite unreliable. More importantly, the results are actually not as clear as they have been made out to be, in two ways. First, there are considerable differences in results regarding the role of the mother's versus the father's carrier status on the sex ratio of the offspring. While in several studies, the elevated sex ratio is related to the mother's HVB carrier status (Chahnzarian *et al.* 1988), in one sample of Melanesian populations, the HVB carrier status of the mother is associated with a significantly *lower* sex ratio at birth (Hesser *et al.* 1976), and while in some studies the father's carrier status has a significant effect on raising the sex ratio at birth, in others it does not (Hesser *et al.* 1976, Chahnazarian *et al.* 1988). In general, the results seem to be weaker on the effect of father's HVB status than on that of mothers, an issue to which I will return below.

Second, the interaction with birth order is ignored. In her Ph.D. dissertation on the sex ratio at birth (the late) Anoush Chahnazarian found that the carrier status of parents appears to increase the sex ratio at birth in low birth orders but not in higher parities (Chahnazarian 1986). She concludes by saying that "the negative relationship observed here between birth order and the sex ratio at birth in children of carrier parents fails to provide an explanation of the unusually high sex ratios at birth observed at higher parities in China, a country of high hepatitis B prevalence" (135).

Thus the empirical evidence from these microstudies is very much open to question. So are the conclusions we can draw from this evidence, such as it is, for the "missing women" countries. Oster reports that there are nine types of hepatitis B virus, with different transmission rates and effects. Given that the link between carrier status and the sex ratio at birth is a pure black-box relationship, it is unclear whether variation in type partly account for the plethora of results in the different microstudies. The type of virus differs by region and none of the microstudies come from the "missing women" countries, so it is unclear which studies reflect the relationship that might prevail there. As a result, it is totally unclear what the evidence from these microstudies suggests about the impact of HVB status on the sex ratio at birth in the "missing women" countries.

Not only is the evidence from these microstudies and its meaning for "missing women" countries inconclusive, but a very large data set from a "missing women" country demonstrates that the HVB carrier status of the mother exerts only a tiny effect on the sex ratio at birth. The study is by Lin and Luoh (2008), who analyze the sex ratio of over 3 million births between 1988 and 1999 to mothers who were included in the Hepatitis B Mass Immunization Program in Taiwan. The very robust result from this study is that there is a positive and significant effect from the HVB status of the mother, but it is extremely small. Using a range of specifications, with or without controlling for covariates, and using different subsamples, the authors find a rather robust effect of the mother's HVB carrier status increasing the likelihood of a child being male by only 0.25 percentage points. Given the prevalence of HVB in China in the 1980s, this would only raise the expected sex ratio at birth by 0.17 percentage points (from 1.06 to 1.0617), with a negligible impact on the calculation of "missing females".

In response to this rather conclusive finding on the relationship between mother's HVB carrier status and the sex ratio at birth, Blumberg and Oster (2007) have recently argued that it is the father's HVB status that is affecting the sex ratio at birth. They take two of the microstudies that are consistent with this view, as well as some aggregate data on HVB prevalence rates of males and females in Taiwan and among Chinese immigrants. In each case, there is some evidence consistent with the view that the HVB status of the father raises the sex ratio at birth. While the Taiwanese data used by Li and Luoh cannot be used to assess this new claim (as the carrier status of the father is not included), this new paper takes a rather selective look at the evidence. The papers that earlier investigated this link using individual microstudies and pooling these studies (including the two used by Blumberg and Oster) found that the link between mother's HVB status and the sex ratio at birth is stronger, while the evidence on father's HVB status is more mixed (e.g. Chahnazarian *et al.* 1988). Also, no interaction with birth order is examined in the new study, so that this new claim has to be viewed with a great deal of skepticism. The need for skepticism is confirmed by the findings from Chen and Oster (2008) who examined the births by HVB status of mothers and fathers in a large sample

in China and came to the conclusion that the HVB carrier status of either parent did not have a significant impact on the sex ratio at birth. As a result Oster now no longer claims that HVB carrier status is a cause of elevated sex ratios in China.

Turning to the time series evidence, Oster (2005) discusses the effect of immunization programs on the sex ratio at birth among populations with high HVB status and finds that immunization campaigns are associated with falling sex ratios at birth. As noted earlier, the case studies she examines are Alaskan Natives and Taiwanese. The discussion above already addresses the issue of the Taiwanese data. Regarding the Alaskan data, these effects are quite small. In the regression which should generate more precise estimates, as it focuses on Native Alaskans (who were the ones with high HVB prevalence), being a Native Alaskan in 1990 leads to a 0.6% increase in the chance of being male, which is rather small indeed and would affect the sex ratio at birth by about one percentage point. Second, being an Alaskan Native has a much larger effect on the sex ratio in whatever time period. This goes against the claim that the unusually high sex ratio at birth among Alaskan Natives is mostly due to their high HVB rates. Third, one should try interactions between Native Alaskans and birth order to see whether the effects are dependent on birth order as well. And as before, it is unclear what this tells us about China and India, given the different types of the virus.

Lastly, Oster's paper presents some cross-country evidence on the link between the sex ratio at birth and HVB prevalence. First, she presents evidence based on WHO classification and shows that the sex ratio at birth is higher in high-prevalence countries, even if one excludes China. The effect is also true within Europe. But the effects are very small. If China is excluded the effect is a little more than one percentage point and would not change the missing women calculation by much. Second, there are no controls for other factors that affect the sex ratio at birth (including overall fertility and mortality conditions). Third, the data on the sex ratio at birth come from the *Demographic Yearbook* and it is not clear whether she uses all of the data, or only those based on 100% registration of births. There are serious problems of under-reporting, including sex-selective under-reporting, and so one would have to be quite careful here.

Oster then collects 63 studies about HVB prevalence from the medical literature and generates prevalence rates by the weighted average of these studies by country. After dropping studies with fewer than 2,500 individuals, she ends up with a sample of 38 countries. A scatterplot is presented for all countries and also one including just OECD countries and then a regression is presented, with the sex ratio at birth on the left-hand side and HVB prevalence and GDP per capita and other controls on the right-hand side. She finds a significant positive influence of HVB prevalence on the sex ratio at birth in all countries (income has no effect, life expectancy a positive effect) and a subsample of OECD countries (16 countries) and in a sample where there is at least 90% complete birth registration.

The scatterplots and the regression have a number of very serious problems, and for several reasons the results cannot be seen as reliable. First, the sample is

very small and it is not clear how representative it is. (In my own analysis with Wink of the influence of life expectancy on the sex ratio at birth, we used more than 200 observations: see Klasen and Wink 2002, 2003). It would be essential to report the HVB rates for all 63 countries (along with sample sizes) to reassure the reader that the sample is not a selective one. Second, the sample is entirely cross-sectional, which raises all sorts of questions about unobserved heterogeneity. The *Demographic Yearbook* does have time series information on sex ratios at birth and one should try to investigate this issue in a panel framework. Third, the reliability of reported sex ratios at birth in the *Demographic Yearbook* is open to question. We know about sex-selective under-reporting of births in China and India (e.g. Johannsson and Nygren 1991; Dyson 2001) and it is likely to be an issue in other countries as well. Fourth, in the figures and the regressions, it appears that the results are heavily influenced by China and South Korea, which ought to be excluded due to questions about the reliability of their reporting of the sex ratio at birth (as well as the impact of sex-selective abortions). Regarding the regressions, one should omit all the countries with evidence of under-reporting of female births and/or evidence of sex-selective abortions. Moreover, the reliability of all the data on the sex ratio at birth need to be checked carefully. Similarly, other factors influencing the sex ratio at birth (e.g. fertility and mortality conditions) should be included in the regressions.

Extending the cross-country analysis, Oster presents evidence on the sex ratios of various immigrant groups in the US, links them to HVB rates in their home countries, and again finds a positive relationship. But again, South Korea, India and China should be thrown out of the analysis for the reasons stated above, including the apparent prevalence of sex-selective abortions. Once this is done, the effect largely disappears. Second, the data concern children, not births, with all the problems stated above with such data. Third, it is bizarre to see that quite a few immigrant groups have sex ratios at birth below 1. Surely this indicates sample size or measurement problems. The regressions should also control for birth order and mortality conditions.

Lastly, Oster discusses a major empirical problem for a link between HVB and the sex ratio at birth, namely evidence from Africa which points to a *lower* (rather than a higher) sex ratio at birth despite high HVB carrier rates (see also Das Gupta 2005). The author claims that there is a positive relationship (results are not shown) and that the low sex ratios at birth in Africa are due to other reasons.

As shown in many studies (including Klasen and Wink 2003), it is indeed the case that Africans have a slightly lower sex ratio at birth. This is well documented in the US, where the difference is about 2 to 3 percentage points (i.e. 1.03 instead of 1.05–1.06); in the Caribbean, the effect is similar (in Africa the data are very sparse but point to similarly lower sex ratios at birth). Thus the race effect is small and, given the high prevalence of HVB in Africa (the rates are among the highest in the world), the presumed HVB effect should easily more than outweigh this race effect.

Thus we are left to wonder how the high prevalence of HVB in Africa does not lead to a higher sex ratio.

While Oster has presented a large amount of evidence, what my discussion has shown is that all of her pieces of evidence suffer from serious problems, while there is considerable evidence suggesting that the link between HVB carrier status and the sex ratio at birth in "missing women" countries is very small or even non-existent. Thus at this stage there is little reason to adjust the estimates of missing women as proposed by Oster. If anything, the evidence seems to suggest that such an effect is very small, with negligible impact on the magnitude of the problem. Oster's retraction of the claim based on her new work in China is further confirmation that the missing women problem cannot be explained away through this link to HVB.

Maybe one of the larger issues regarding this question is the sheer size of the estimates of missing women, which may strike some observers as simply too large to be the result of gender bias in mortality. But a back-of-the-envelope calculation might help to clarify how these numbers can come about. First, it has to be remembered that the number of missing women is a stock, not a flow, thus representing the impact of past and present (pre- and post-birth) gender bias in mortality on the sex composition of the population at the census date. How could India's 38 million and China's 34 million missing females in 1991 have come about? Assume that all gender bias in mortality took place among children below 5 and this gender bias ensured that under-5 mortality rates for females were some 50% higher than they would have been in the absence of gender bias in mortality. For the generations alive in 1991, the average female under-5 mortality rate might have been 150 per 1,000 instead of 100 per 1,000 in India (and 90 instead of 60 per 1,000 in China). Given about 20 million births on average among the cohorts alive today in both India and China, that would generate about 1 million missing females per year in India and 0.6 million in China. Aggregation over the cohorts (and allowing for lower life expectancy in India than in China among all the cohorts) could easily generate the numbers of missing females found for the early 1990s. The increasing spread of sex-selective abortions can then have contributed to the increase of the phenomenon, particularly in China. This rough and ready calculation can only give a rough impression of how these staggeringly large numbers might have come about.[12]

IV. Conclusion

This short essay has focused on recent discussions of the role of fertility decline and the Hepatitis B carrier status on levels and trends in gender bias in mortality. The

[12] A more careful assessment would have to consider birth and mortality rates over time by sex and cohort; given data limitations, this cannot easily be accomplished.

good news is that there is little reason to worry that a fertility decline occurring as a response to overall economic and social development will intensify gender bias in mortality. Only forced fertility decline through coercive family planning policies is likely to have such an effect. The bad news is that the evidence presented by Oster claiming to have found about 45% of the "missing females" is weak and contradicted by more powerful evidence to the contrary. Thus it still is the case that gender bias in mortality is as large as presumed by the calculations made by Sen, Coale, and Klasen and Wink. But even in the bad news there is some good news. While the implication of Oster's claims had been that gender bias in mortality has been getting much worse recently, the fact that her claim appears to rest on weak foundations suggests that the earlier suggestions made by Klasen and Wink (2002, 2003) still seem to be valid, namely that in most parts of the "missing women" regions, there has been a slight to moderate improvement. The exception is China, where the one-child policy, combined with the increasing availability of sex-selective abortions, has led to lower female survival.

References

Abrevaya, J. (2005), "Are There Missing Girls in the United States?", unpubl. paper, Purdue University.

Attane, I. (2007), "Female Discrimination in China: Provincial Trends and Determinants", unpubl. paper, INED, Paris.

Banister, J. (1987), *China's Changing Population* (Stanford, Calif.: Stanford University Press).

—— and Coale, A. "Five Decades of Missing Females in China", *Demography*, 31: 459–79.

Bardhan, P. (1974), "On Life and Death Questions", *Economic and Political Weekly*, 9: 1283–304.

Barro, R. (2005), "The Case of Asia's Missing Women", *Business Week*, 28 February 12.

Basu, A. (1999), "Fertility Decline and Increasing Gender Imbalance in India, Including a Possible South Indian Turnaround", *Development and Change*, 30: 237–63.

—— (2000), "Fertility Decline and Worsening Gender Bias in India: A response to S. Irudaya Rajan et al.", *Development and Change*, 21: 1093–5.

Bhattacharya, P. (2006), "Economic Development, Gender Inequality, and Demographic Outcomes: Evidence from India", *Population and Development Review*, 32: 263–91.

Blumberg, B., and Oster. E. (2007), "Hepatitis B and Sex Ratios at Birth: Fathers or Mothers?", unpubl. paper, University of Chicago.

Chahnazarian, A. (1986), "Determinants of the Sex Ratio at Birth", unpubl. Ph.D. diss., Princeton University.

—— Blumberg, B., and London, W. Th. (1988), "Hepatitis B and the Sex Ratio at Birth: A Comparative Study of Four Populations", *Journal of Biosocial Sciences*, 20: 357–70.

Chen, G., and Oster, E. (2008), *Hepatitis B Does Not Explain Male-Biased Sex Ratio in China*, NBER Working Paper 13971 (Cambridge, Mass.: National Bureau of Economic Research).

Chung, W., and Das Gupta, M. (2007), "Why is Son Preference Declining in South Korea?", *Population and Development Review*, 33(4): 757–83.

Coale, A. (1991), "Excess Female Mortality and the Balance of the Sexes: An Estimate of the Number of Missing Females", *Population and Development Review*, 17: 517–23.

Das Gupta, M. (2005), "Explaining Asia's Missing Women: A New Look at the Data", *Population and Development Review*, 31: 539–35.

—— (2006), "Cultural versus Biological Factors in Explaining Asia's Missing Women: Response to Oster", *Population and Development Review*, 32: 328–32.

—— and Mari Bhat, P. N. (1997), "Fertility Decline and Increased Manifestation of Sex Bias in India", *Population Studies*, 51: 307–15.

Drèze, J., and Murthi, M. (2001), "Fertility, Education and Development: Evidence from India", *Population and Development Review*, 27: 33–63.

—— and Sen, A. (2001), *India: Development and Participation* (New York: Oxford University Press).

Dyson, T. (2001), "The Preliminary Demography of the 2001 Census in India", *Population and Development Review*, 27: 257–70.

Ebenstein, Avraham (2007), "Fertility Choices and Sex Selection in Asia: Analysis and Policy", unpubl. paper, University of California, Berkeley.

Hesser, J. E., Blumberg, B. S., and Drew, J. S. (1976), "Hepatitis B. Surface Antigen, Fertility, and Sex Ratio: Implications for Health Planning", *Human Biology*, 48: 73–81.

Johansson, S., and Nygren, O. (1991), "The Missing Girls of China: A New Demographic Account", *Population and Development Review*, 17: 35–51.

Klasen, S. (1994), "Missing Women Reconsidered", *World Development*, 22: 1061–71.

—— (2003), "Sex Selection", in P. Demeny and G. McNicoll (eds), *Encyclopaedia of Population* (New York: Macmillan).

—— and Wink, C. (2002), "A Turning Point in Gender Bias in Mortality: An Update on the Number of Missing Women", *Population and Development Review*, 28: 285–312.

—— —— (2003), "Missing Women: Revisiting the Debate", *Feminist Economics*, 9: 263–99.

Lai, D. (2005), "Sex Ratio at Birth and Infant Mortality Rate in China: An Empirical Study", *Social Indicators Research*, 70: 313–26.

Lin, M.-J., and Luoh, M.-C. (2008), "Can Hepatitis B Mothers Account for the Number of Missing Women? Evidence from 3 Million Newborns in Taiwan", American Economic Review (forthcoming).

Mari Bhat, P. N., and Francis Zavier, A. J. (2003), "Fertility Decline and Gender Bias in Northern India", *Demography*, 40: 637–57.

Murthi, M., Guio, A.-C., and Drèze J. (1995), "Mortality, Fertility, and Gender Bias in India: A District-Level Analysis", *Population and Development Review*, 21: 745–82.

OECD (2006), *Gender, Institutions, and Development Database* (Paris: OECD).

Oster, E. (2005), "Hepatitis B and the Case of Missing Women", *Journal of Political Economy*, 113: 1163–216.

Rajan, S. Irudaya, Sudha, S., and Mohanachandran, P. (2000), "Fertility Decline and Worsening Gender Bias in India: Is Kerala no Longer an Exception?", *Development and Change*, 31: 1085–95.

Registrar General (2003), *Final Population Totals, Census of India 2001* (New Delhi: Ministry of Home Affairs).

Sen, A. (1989), "Women's Survival as a Development Problem", *Bulletin of the American Academy of Arts and Sciences*, 43(2): 14–29.

—— (1990a), "Gender and Cooperative Conflict", in I. Tinker (ed.), *Persistent Inequalities* (New York: Oxford University Press).

—— (1990b), "More Than 100 Million Women are Missing", *New York Review of Books*, 20 December.

—— (1992), "Missing Women", *British Medical Journal*, 304: 587–8.

—— (1998), *Development as Freedom*, (New York: Knopf).

—— and SENGUPTA, S. "Malnutrition of Rural Children and the Sex Bias", *Economic and Political Weekly*, 18: 855–64.

VISARIA, P. (1961), *The Sex Ratio of the Population of India*, Census of India 1961, Monograph 10 (New Delhi: Office of the Registrar General).

World Bank (2007), *World Development Indicators* (Washington, DC: World Bank).

PART III

GROWTH, POVERTY AND POLICY

This page intentionally left blank

CHAPTER 16

CHALLENGES OF ECONOMIC DEVELOPMENT IN PUNJAB

ISHER AHLUWALIA

This chapter addresses the challenges of economic development facing Punjab. In a ranking based on the Human Development Index (HDI), which combines literacy rate, life expectancy and per capita income in a weighted index, Punjab ranks second only to Kerala among 15 major states of India (Figure 16.1). It also fares much better than Haryana and remains ahead of Maharashtra and Gujarat, the three states that have overtaken it in per capita income in recent years. Since the focus of policy-makers in India is largely on poverty alleviation in the poorest states, it is implicitly believed that states such as Punjab, which have largely eliminated income poverty, are on track to achieving sustainable increases in the quality of life of their residents. Yet, despite its high income, low income-poverty, and high HDI ranking, Punjab faces serious challenges on economic growth as well as on the human development front, the latter arising principally from weak and malfunctioning public systems of education and health, and a deplorable record on gender discrimination.

I am thankful to Anurag Aggarwal, Montek Ahluwalia, Ravi Kanbur, Shiva Kumar, Vrinda Sarup, Amarjeet Sinha, T. Sundaraman and Michael Walton for their very helpful comments. My thanks to Annu Gupta for her research assistance and Deepa Gopalan for providing secretarial support.

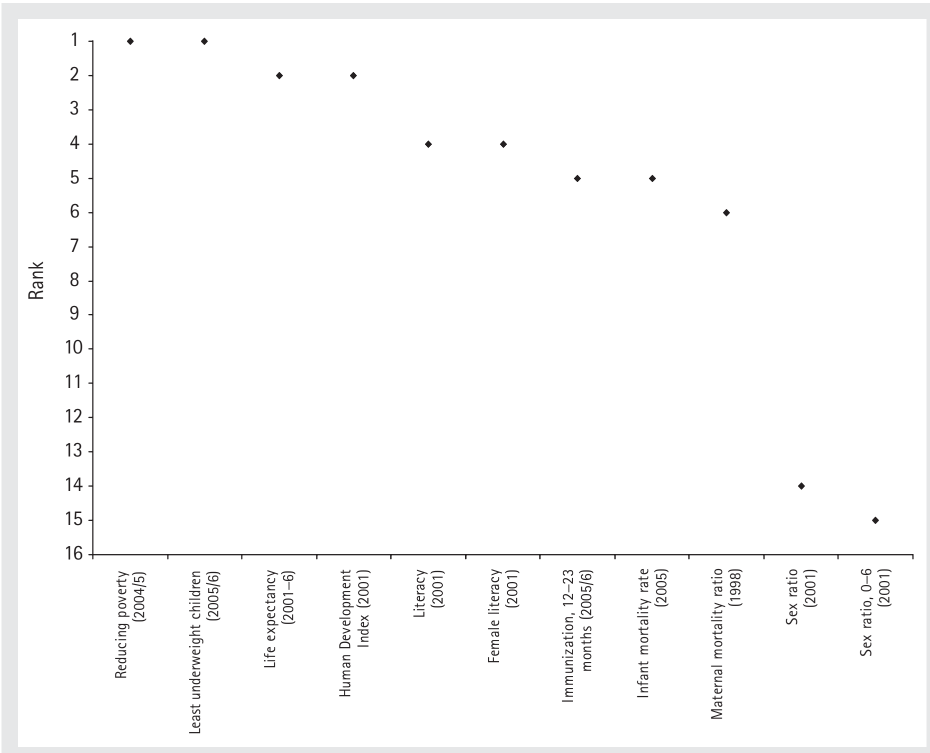


Fig. 16.1. Punjab's rank among 15 major states of India on various social indicators

Punjab has suffered a setback in economic growth over the past two decades, partly because of the militant political movement in the state during the 1980s, but more importantly because government policies have not been able to stimulate growth even after the end of militancy in the early 1990s. The prospects for sustained economic growth driven by agriculture—the traditional economic engine of the state—are severely limited by the erosion of natural resources. An important part of the solution of reviving agricultural growth will have to come from an industrial strategy which builds synergies with the agricultural sector. Preoccupation with agriculture has meant that not enough attention has been paid to developing a sound industrial strategy for the state. The government administration, relatively inefficient compared with other states, has resulted in a high transactions cost of doing business, which has adversely affected the flow of private investment into industry in the state. Rising fiscal deficits have constrained the ability of the state to spend more on health and education, while the deterioration in governance has weakened the public systems of health and education. With faltering growth and limited ability to develop human capital, the challenges of economic development are formidable.

Section I presents Punjab's record on human resource development. Section II spells out its economic growth experience. Section III presents the issues with regard to the sustainability of the high growth rates of the past, particularly the environmental challenge. It also highlights the limits imposed by high levels of fiscal deficits on addressing the challenge of accelerating growth, and the poor level of delivery of satisfactory public services.

I. Human Resource Development

Planners in India have long been designing schemes for alleviating poverty. In recent years, inspired by the writings of Amartya Sen and work at the United Nations Development Programme (UNDP), other dimensions of development, including the gender dimension, have received more attention. The Tenth Five-Year Plan of India (2002/3 to 2006/7) was the first to specify measurable targets for specific indicators of development. It specified progress on these indicators at a faster pace than the eight Millennium Development Goals (MDGs) adopted by India as a signatory to the Millennium Declaration in 2000/1. In recent years, the government of India has restructured and rationalized many of its schemes and emphasized decentralization in their execution and those of other new programs which are explicitly targeted to promote human resource development. The government of Punjab, like other state governments, is also emphasizing the need to improve the delivery of education and health services.

I.1 Progress on Millennium Development Goals (MDGs)

The MDGs provide a useful framework within which to review progress in development in Punjab.

Eradicating Extreme Poverty and Hunger (MDG1)

Punjab has the lowest incidence of poverty of any state in India. Official estimates of poverty, based on the cost of an all-India average consumption basket at which calorie norms are met,[1] place 8.4% of the population of Punjab in poverty in 2004/5 (down from 11.8% in 1993/4), compared with the all-India average of 27.5%.

[1] The norms are 2,400 calories per capita per day for rural areas and 2,100 calories for urban areas. As of 1999/2000, the calorie norms, when expressed in monetary terms and adjusted for inflation, amounted to Rs327 per month per capita for rural areas and Rs454 per month per capita for urban areas. These estimates are based on a uniform reference period.

Table 16.1. Poverty, Punjab and all India (headcount ratio, %)

	Punjab	All India
1973/4	28.2	54.9
1977/8	19.3	51.3
1987/8	13.2	38.9
1993/4	11.8	36.0
1999/2000	not available (6.2)	not available (26.1)
2004/5	8.4 (5.2)	27.5 (21.8)

Notes: Poverty headcount ratio is defined as the percentage of the population below the poverty line. Data are based on the Uniform Reference Period (URP). Figures in parentheses for 1999/2000 and 2004/5 are estimates based on the Mixed Reference Period (MRP).
Source: Planning Commission.

A parallel set of "official" estimates based on the Mixed Reference Period (MRP) places poverty in Punjab at 5% in 2004/5. Alternative estimates prepared by Deaton and Drèze (2002) suggest the proportion to be even lower, at 3% in 1999/2000, compared with 6.7% in 1993/4.

This reduction in poverty has been achieved through rapid economic growth in the rural sector, while good initial conditions in physical infrastructure were also helpful. The maximum reduction in Punjab, 8.9 percentage points, was between 1973 and 1978, when Punjab led the country in the Green Revolution in agriculture (Table 16.1). Poverty in Punjab is concentrated in the socially disadvantaged groups, e.g. the scheduled castes, who account for 71% of the poor, compared with 36% on average in other states.[2]

As for eradicating hunger and malnutrition, the MDG for India is to halve the proportion of underweight children (reflecting acute and chronic malnutrition) from 51.5% in 1992/3 to 25.7% in 2015.[3] Punjab's record on child nutrition during the 1990s was impressive in contrast with that of most other states (Table 16.2). At 27% in 2005/6, Punjab now enjoys the distinction of having the smallest proportion of children who are underweight. But getting from 46% in 1992/3 to 23% by 2015 will require focused effort because progress has slackened considerably on this front since 1998/9 in Punjab, as in the rest of the country.

The Integrated Child Development Services program (ICDS), the main government program on child nutrition in India since 1975, has played a relatively smaller

[2] Punjab has the highest population of scheduled castes of any state in India (28.9%). See Radhakrishna *et al.* (2007).

[3] As shown in Table 16.2, the data on undernutrition based on the new 2006 WHO standards put the proportion of underweight children in India at 42.5% and in Punjab at 24.9%. These data, however, cannot be compared with the data from the earlier National Family Health Surveys, which are based on US standards.

Table 16.2. Underweight children (below 3 years of age), Punjab and other Indian states

State	1992/3		1998/9		2005/6		2005/6[a]	
	%	Rank	%	Rank	%	Rank	%	(WHO) Rank
Punjab	46.0	6	28.7	2	27.0	1	24.9	2
Andhra Pradesh	45.0	4	37.7	6	36.5	4	32.5	4
Assam	49.2	8	36.0	4	40.4	7	36.4	5
Bihar	62.6[b]	15	54.3	15	58.4	16	55.9	16
Chattisgarh	nr	nr	60.8	18	52.1	15	47.1	15
Gujarat	48.1	7	45.1	9	47.4	14	44.6	14
Haryana	34.6	2	34.6	3	41.9	9	39.6	10
Jharkhand	nr	nr	54.3	15	59.2	17	56.5	17
Karnataka	50.6	9	43.9	8	41.1	8	37.6	7
Kerala	27.0	1	26.9	1	28.8	2	22.9	1
Madhya Pradesh	57.4[b]	13	53.5	14	60.3	18	60.0	18
Maharashtra	51.4	10	49.6	11	39.7	6	37.0	6
Orissa	52.4	11	54.4	17	44.0	11	40.7	12
Rajasthan	44.3	3	50.6	12	44.0	11	39.9	11
Tamil Nadu	45.7	5	36.7	5	33.2	3	29.8	3
Uttaranchal	nr	nr	41.8	7	38.0	5	38.0	8
Uttar Pradesh	59.0[b]	14	51.8	13	47.3	13	42.4	13
West Bengal	54.8	12	48.7	10	43.5	10	38.7	9
All India	51.5		47.0		45.9		42.5	

Note: "nr" = "not relevant", as Chattisgarh, Jharkhand and Uttaranchal were created out of Madhya Pradesh, Bihar and Uttar Pradesh respectively in the year 2000.

[a] Based on the new WHO standards for children below 5 years of age, which cannot be compared with the earlier data from National Family Health Surveys, based on US standards.

[b] Data are for children below 4 years of age.

Source: NFHS (1992/3, 1998/9 and 2005/6).

role in Punjab, since high per capita income ensures good nutrition at home. Better nutrition for women has also helped Punjab to attain better nutrition levels for children than in other states.

Achieving Universal Primary Education (MDG2)

The Millennium Development Goal with respect to primary education is to ensure that by 2015 all children are able to complete primary education. Punjab has yet to achieve this goal, though it is well placed to do so. Punjab performs extremely poorly in quality of primary education.

The educational infrastructure in Punjab is among the best in India. By 2001, 96% of habitations in Punjab had a primary school within a distance of one

kilometre.[4] As of 2005/6, 93% of all schools were equipped with drinking water facilities, 78% had common toilet facilities and 70% were equipped with toilets for girls. However, this infrastructure advantage is offset by Punjab's poor score on teachers and outcomes, so much so that the recently published Educational Development Index for all primary schools under government and government-aided management ranks Punjab 14th among the 35 states and union territories of India (HRD 2007).

It is clear that when the main need of the education and health sectors was setting up physical infrastructure through the building of schools, hospitals and roads, political support was easier to procure through patronage in construction activity, a phenomenon now being experienced by many states which are currently setting up their infrastructure for education and health. The relative failure of the system lies in the area of public action to ensure effective delivery of education. Alignment of these policies with political structures has been a much more challenging task.

Three nationwide surveys on learning achievement of students conducted by the National Council of Educational Research and Training (NCERT) in 2002 and 2004 revealed that Punjab is consistently below the national average, with the exception of Maths for class V students (NCERT 2005, 2008). Indeed, as Figure 16.2 shows, the results for Punjab students are among the worst. The language achievement results for class VII/VIII are presented against per capita income in Figure 16.3a, which highlights a number of deficits in social development in Punjab relative to income.

Absenteeism of teachers in government schools is a major factor contributing to the poor achievements in learning in Punjab. Kremer *et al.* (2005) present strong evidence of teacher absenteeism in their survey of 2002/3 which ranks Punjab 17th among 19 Indian states, and finds that more than 34% of teachers were absent during unannounced visits to government primary schools. The latest ASER Survey also shows that teacher absenteeism in Punjab in 2005 was the third highest after West Bengal and Haryana. However, the ASER Survey for the year 2007 shows a distinct improvement, reflecting the effect of the hiring of 12,000 teachers by the PRIs in 2006. Even so, the teacher absenteeism at 14.7% in Punjab in 2007 was the highest of all states, and much higher than the national average of 9%.

Another factor contributing to the poor outcomes of teaching in Punjab is the inadequate attention paid to the development of course content. The State Council of Education, Research and Training (SCERT) is perhaps a more appropriate institution than the Punjab School Education Board for developing and evolving the curriculum for primary schools. An important initiative of the government of Punjab in this respect has been to promote the use of information and

[4] The greatest expansion of schooling facilities took place in the 1970s, when the number of schools in Punjab increased by two-thirds. For more information, see UNDP (2004: 102–3).

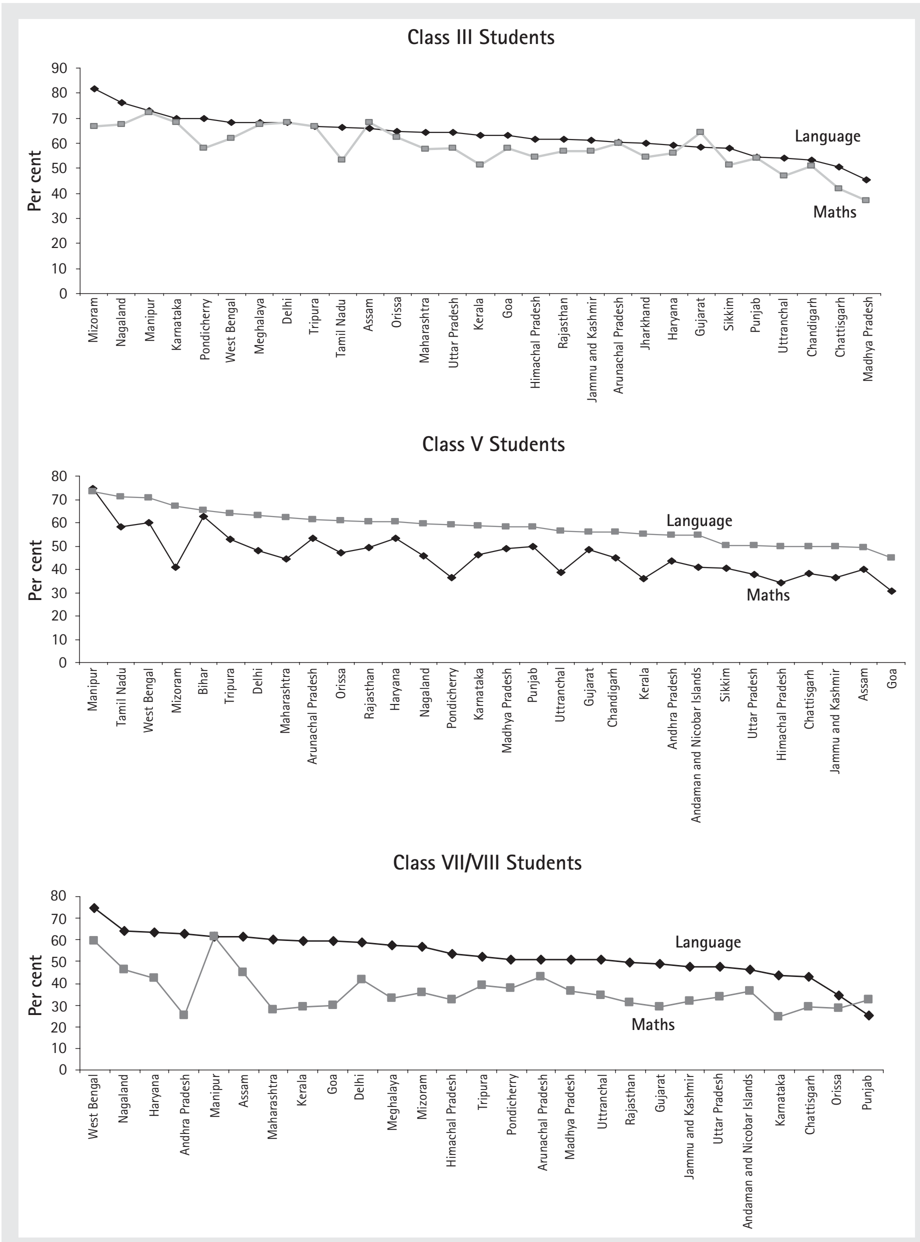


Fig. 16.2. Learning achievement in India: language and maths

Notes: The survey for class V was conducted in 2002 and the report was released in 2005. The other two surveys were conducted in 2004; these data were released in 2006 and the reports are expected to be released in July 2008.

Source: Department of Educational Measurement and Evaluation, National Council of Educational Research and Training (NCERT).

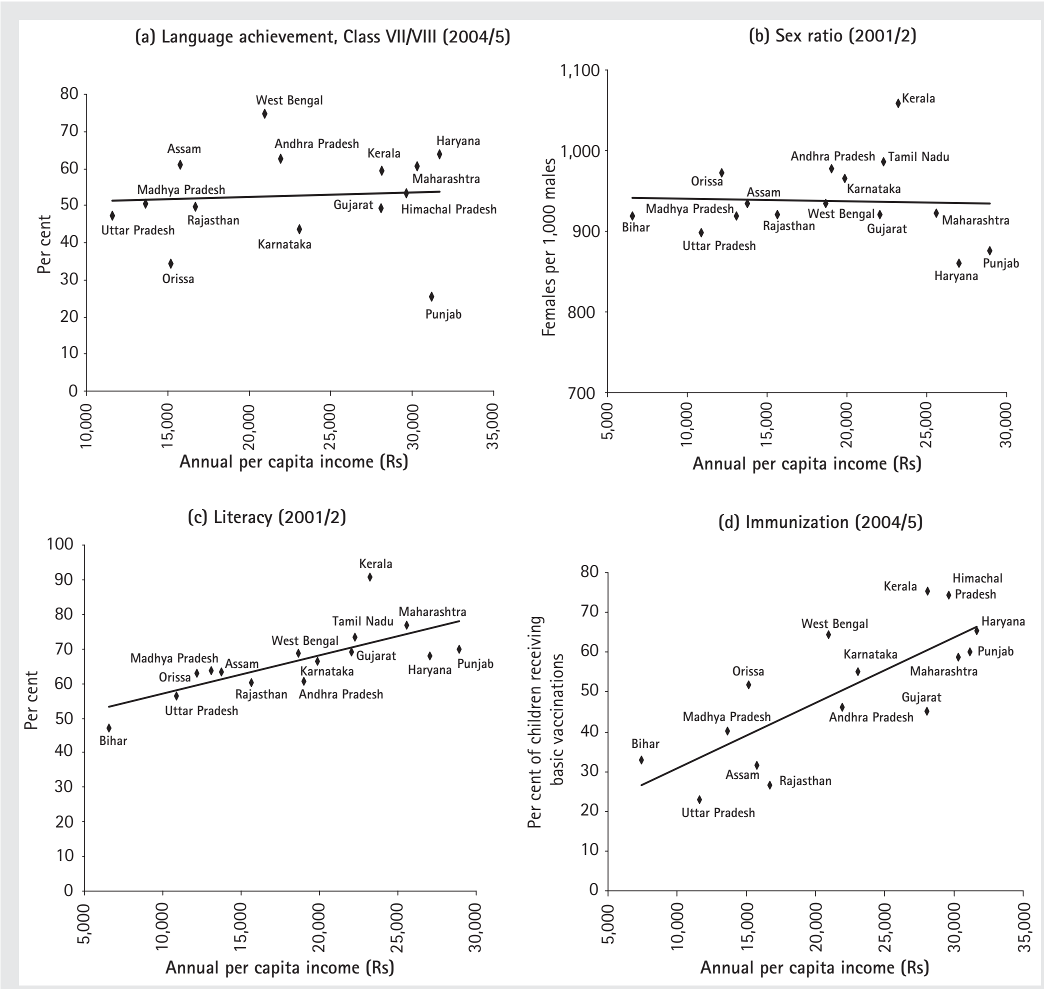


Fig. 16.3. Education, gender and health deficits, Punjab

communications technology (ICT) in developing and delivering multimedia curriculum content for teaching in upper primary schools. The ICT project, begun in 2005, holds out tremendous promise if two major challenges can be overcome, i.e. availability of electricity and training of teachers.[5]

A centralized system of hiring of teachers at very high salaries (compared with private schools), combined with political interference at the state level, has meant that patronage and corruption have had a free hand in hiring. Moreover, very poor supervision and accountability to remote authorities has encouraged those appointed in rural schools who live in urban and/or distant locations to collect

[5] Computer training of teachers and monitoring of the ICT project is currently being done in 600 schools through a memorandum of understanding with the America India Foundation under the Digital Equalizer Program.

their salary as teachers while informally subcontracting the job at a discounted salary to other, less qualified or unqualified teachers. The weak incentive structure for promotions and an ineffective supervisory framework encourages a high rate of teacher absence.

The rising demand for good education has led to a mushrooming of private, unrecognized fee-paying schools in Punjab, including the rural areas, which are showing a rapid increase in enrollment. At the same time, enrollment in government schools in Punjab has been declining for some time, and the actual decline is much larger than suggested by the official data.[6] A recent, officially sponsored study of the unrecognized schools in seven districts of Punjab by Mehta (2005) finds that enrollment in these schools accounts for 25.8% of the total enrollment in primary schools. Mehta finds that the unrecognized schools have more favorable pupil–teacher ratios,[7] a greater percentage of female teachers, better-qualified teachers than in recognized schools, but less training and no provision for in-service training. The study also finds that more of these schools use English as the medium of instruction, and their facilities and general infrastructure are also better than in recognized schools. The unrecognized schools have not opted for "recognition", either because they find the administrative procedures of recognition very cumbersome or because they are reluctant to open their books to inspectors and/or authorities.

In their nationwide survey, Muralidharan and Kremer (forthcoming 2008) also find that in 2003 Punjab had one of the highest percentages of villages with a private school, 65% compared with an all-India average of 28%. Their study shows a high positive correlation between absenteeism among public school teachers and the prevalence of private schools.

Since those who are willing and able to pay to escape the poor quality of education in government schools can send their children to unrecognized schools, it is poor children who suffer the most. It is worth investigating whether the Mid-Day Meal Scheme is being enforced in government schools and helping to stem the outflow, or whether the quality deficit is so high as to continue to drive students to private schools.

Punjab is one of the few states not covered under the long-standing DPE program of the government of India, designed to improve service delivery by government schools. As a result, the state has lagged behind in institutional structures for teacher training and support. It has also not devolved educational responsibilities to the Panchayati Raj Institutions (PRIs), unlike Maharashtra and Gujarat and more recently Madhya Pradesh, Karnataka, Bihar, Uttar Pradesh and Rajasthan. As of September 2007, only 5,500 of the 18,000 schools were transferred to the Panchayati Raj Institutions with authority for local recruitment of teachers. Mostly, local communities do not have power over hiring, promotions and imposition of

[6] Official data indicate a decline in enrollment from 84.4% in 2005/6 to 82% in 2006/7 for primary schools and from 84% to 81.7% for upper primary schools.

[7] As regards the teacher–pupil ratio in government schools, it is affected by the government's ability to fill vacancies, on the one hand, and declining enrollment, on the other.

penalties on teachers who are absent and not teaching. Essentially, Punjab has not yet trusted local bodies to handle elementary education and primary health. This is ironic, since Punjab has been one of the foremost states to structure its community development through Panchayati Raj.

The launch of the Sarva Shiksha Abhiyan (SSA), the campaign for universal primary education by the government of India in 2001, brought Punjab into the fold of the national program. The proceeds from the 2% levy on all income by the government of India are earmarked for elementary education. The SSA initiative attempts to fill the gaps in the state government education infrastructure. It has led to greater emphasis on design, delivery and evaluation of education at the state level within a decentralized framework. States like Punjab, however, are at a disadvantage because the allocation of funds is typically linked to inadequacy of physical infrastructure, while Punjab requires funds from the government of India for teacher training, curriculum development, etc.

Efforts by the state government to bring private management into the running of government schools a few years ago were frustrated by vested interests. A new initiative by the current government to get the private sector to set up Adarsh schools also misses the point, in that the problem with the schools is not with infrastructure but with the management structures. The challenge is to improve overall governance by designing a system that delivers education of good quality and is accountable to users. This will require the elimination of political interference in the management of government schools.

Promoting Gender Equality and Empowering Women (MDG3)

Punjab's experience shows that unequal gender relations tend to be independent of economic growth, and there is a tremendous need for public action in closing the gender gap. Punjab's severe sex-ratio deficit can be seen in Figure 16.3b, which presents sex ratio mapped against per capita income for the 15 major states of India. While sex ratio is not correlated with the level of per capita income, it is much lower in Punjab than the average of all states.

Gender disparities in Punjab begin before birth due to the abhorrent practice of female feticide. The traditional preference for sons is driven by an intensely patriarchal mindset and attachment to land. Punjab is notorious, together with states such as Haryana, Rajasthan and Gujarat, for a long-standing practice of "eliminating" the girl child (the Kurimar tradition). This practice has evolved in recent years to a practice of female feticide thanks to ultrasound technology. Punjab has the second lowest (female–male) sex ratio in India (876), and one of the worst in the world (Table 16.3a). All of the 17 districts of Punjab were amongst the 34 districts of the entire country where this sex ratio was the poorest. The child sex ratio in Punjab declined from 875 in 1991 to a low of 798 in 2001 (Table 16.3b).

Table 16.3. Sex ratio, Punjab and other Indian states

State	1981		1991		2001	
	Ratio	Rank	Ratio	Rank	Ratio	Rank
(a) All ages						
Punjab	879	13	882	13	876	14
Andhra Pradesh	975	4	972	3	978	3
Assam[a]	–	–	923	9	935	6
Bihar	946	6	911	11	919	11
Gujarat	942	7	934	6	920	10
Haryana	870	14	865	15	861	15
Karnataka	963	5	960	5	965	5
Kerala	1,032	1	1,036	1	1,058	1
Madhya Pradesh	941	8	931	8	919	11
Maharashtra	937	9	934	6	922	8
Orissa	981	2	971	4	972	4
Rajasthan	919	10	910	12	921	9
Tamil Nadu	977	3	974	2	987	2
Uttar Pradesh	885	12	879	14	898	13
West Bengal	911	11	917	10	934	7
All India	934		927		933	
(b) 0–6 years						
Punjab	908	13	875	15	798	15
Andhra Pradesh	992	2	975	1	961	2
Assam[a]	–	–	975	1	965	1
Bihar	981	3	953	7	942	7
Gujarat	950	11	928	11	883	13
Haryana	902	14	879	14	819	14
Karnataka	974	6	960	5	946	6
Kerala	970	7	958	6	960	3
Madhya Pradesh	977	5	941	10	932	9
Maharashtra	956	9	945	9	913	11
Orissa	995	1	967	3	953	5
Rajasthan	954	10	916	13	909	12
Tamil Nadu	967	8	948	8	942	7
Uttar Pradesh	935	12	927	12	916	10
West Bengal	981	3	967	3	960	3
All India	962		945		927	

[a] No census was held in Assam in 1981.
Source: Census of India.

Of the ten districts in India that have the lowest child sex ratio, seven are from Punjab (UNDP 2004).

The Pre-Natal Diagnostic Techniques Act of 1994 was supposed to prevent misuse of technology in facilitating female feticide, but enforcement has been very weak. Incentive schemes such as the Balri Rakshak Yojna, under which the state

government provides incentives for adopting a terminal method of sterilization after the birth of only one or two girl children, and schemes under which a village is rewarded by a grant if it achieves a sex ratio of 951 or above, are combined with an awareness-raising campaign via print and electronic media. But there is need for greater involvement of civil society in changing long-held and culturally shaped social preferences.

There is evidence of a turnaround in a few districts. A particularly noteworthy example is that of Nawanshahr, where the child sex ratio was as low as 774 in March 2005. By March 2006 the ratio had exceeded 900 in 77 out of the 475 villages of the district, thanks to a vigorous social audit of pregnant mothers, a vigorous medical audit of scanning centers, and the active involvement of NGOs, religious leaders and young students. Of course, sustaining this turnaround will be a major challenge.

As for the gender gap in enrollment for primary education in government schools, Punjab appears not to be doing too badly compared with the national average. However, there is evidence of gender bias in access to the fast-growing, unrecognized, fee-paying private schools. Mehta (2005) finds that for every 100 boys enrolled, these are only 68 girls. Punjab made some progress in closing the gender gap in female literacy during the 1990s, and at 63.4% is ahead of the national average of 53.7% (Table 16.4), but this achievement is not so impressive when measured against per capita income (Figure 16.3c).

Table 16.4. Literacy rate, Punjab and other Indian states (%)

State	1981		1991		2001	
	Male	Female	Male	Female	Male	Female
Punjab	55.6	39.6	65.7	50.4	75.2	63.4
Andhra Pradesh	46.8	24.2	55.1	32.7	70.3	50.4
Assam[a]	–	–	61.9	43.0	71.3	54.6
Bihar	46.6	16.5	52.5	22.9	59.7	33.1
Gujarat	65.1	38.5	73.1	48.6	79.7	57.8
Haryana	58.5	26.9	69.1	40.5	78.5	55.7
Karnataka	58.7	33.2	67.3	44.3	76.1	56.9
Kerala	87.7	75.7	93.6	86.2	94.2	87.7
Madhya Pradesh	48.4	24.0	58.4	28.9	76.1	50.3
Maharashtra	69.7	41.0	76.6	52.3	86.0	67.0
Orissa	56.5	25.1	63.1	34.7	75.4	50.5
Rajasthan	44.8	14.0	55.0	20.4	75.7	43.9
Tamil Nadu	68.1	40.4	73.8	51.3	82.4	64.4
Uttar Pradesh	47.5	17.2	55.7	25.3	68.8	42.2
West Bengal	59.9	36.1	67.8	46.6	77.0	59.6
All India	56.4	29.8	64.1	39.3	75.3	53.7

[a] Census was not held in Assam in 1981.
Source: Census of India.

Improving Health Indicators (MDG4, MDG5, MDG6)

The Millennium Development Goals for improving health conditions focus on infant mortality, immunization and maternal mortality, on the one hand, and diseases such as tuberculosis, malaria and the increasing challenge of HIV-AIDS on the other. Not only is the overall health picture in Punjab incommensurate with its relatively high income level but there has also been a distinct deterioration in the more recent period, even though Punjab increased its spending on health from 0.9% of GSDP in 1995/6 to 1% in 2004/5 and its government spends more on health per capita than the average of the 15 major states (Rao and Choudhury 2008).

In immunization by basic vaccinations, Punjab was well ahead of the all-India average of 35.4% in 1992/3 and 42% in 1998/9, with 62% of its children covered in 1992/3 and 72% in 1998/9 (Table 16.5). As a nationwide campaign for the pulse polio vaccination gathered momentum in recent years, Punjab as well as many of the other high-coverage states, such as Tamil Nadu, Maharashtra, and Madhya Pradesh, neglected their basic vaccinations. Punjab's social deficit conditional on per capita income, with a coverage of basic vaccinations of only 60%, can be seen in Figure 16.3d.

Reliable data on maternal mortality are available only for recent years (Table 16.5). Punjab's relatively poor performance (sixth among the 15 states) reflects the absence of adequate support for deliveries, showing that timely medical support can literally make the difference between life and death. Both in births attended by skilled health personnel and in the percentage of births delivered at institutions, Punjab ranks in the middle of the 15 major states (sixth and seventh, respectively).

The infant mortality rate (IMR) in Punjab declined significantly during the 1980s, but there was no change between 1991 and 1999 (Table 16.5), notwithstanding the relatively high rate of female literacy and high per capita income. The period from 1999 to 2005 saw a decline in IMR from 53 to 44. Much of the infant mortality in Punjab, as in the rest of the country, is concentrated in high neo-natal mortality.

As regards communicable diseases, Punjab has the lowest incidence of tuberculosis of all states except Karnataka, and a 100% cure rate for those detected (Table 16.6). Its incidence of HIV-AIDS is 12th out of 18 states, with 0.12% of its adult population affected. The incidence of malaria in Punjab, however, is much higher than in Kerala, Tamil Nadu and Karnataka, and reflects poor environmental hygiene.

A recent study by Paul *et al.* (2004), based on a stratified random sample of households across the states of India, provides a comparative perspective on access to, use and reliability of, and satisfaction with government health services. While ease of access to government health care facilities for rural areas in Punjab is the highest among the major states (62%), it is much lower than the national average for urban areas (24% vs 41%). Punjab does surprisingly well on the reliability/presence of doctors and paramedics, but only 2–3% of patients express "full satisfaction" with

Table 16.5. Health Indicators, Punjab and other Indian states

State	Infant mortality rate (IMR)								Immunization by basic vaccinations (IMMB)						Maternal mortality ratio (MMR)[a]					
	1981		1991		1999		2005		1992/3		1998/9		2005/6		1997/8		1999–2001		2001–03	
	IMR	Rank	IMR	Rank	IMR	Rank	IMR	Rank	IMMB	Rank	IMMB	Rank	IMMB	Rank	MMR	Rank	MMR	Rank	MMR	Rank
Punjab	81	4	53	2	53	6	44	6	61.9	3	72.1	4	60.1	5	280	8	177	5	178	6
Andhra Pradesh	86	5	73	9	66	10	57	10	45.4	8	58.7	7	46.0	11	197	6	220	8	195	8
Assam	106	9	81	12	76	13	68	14	19.4	14	17.0	16	31.4	16	568	14	398	10	490	14
Bihar	118	12	69	6	63	8	61	12	10.7	15	11.6	17	32.8	15	531	13	400	11	371	11
Chattisgarh	nr	nr	nr	nr	78	14	63	13	nr	nr	21.8	13	48.7	10	nr	nr	nr	nr	nr	nr
Gujarat	116	11	69	6	63	8	54	9	49.8	7	53.0	8	45.2	12	46	1	202	6	172	5
Haryana	101	8	68	5	68	11	60	11	53.5	5	62.7	5	65.3	3	136	3	176	4	162	4
Jharkhand	nr	nr	nr	nr	71	12	50	7	nr	nr	8.8	18	34.5	14	nr	nr	nr	nr	nr	nr
Karnataka	69	2	77	10	58	7	50	7	52.2	6	60.0	6	55.0	8	245	7	266	9	228	9
Kerala	37	1	16	1	14	1	14	1	54.4	4	79.7	2	75.3	2	150	4	149	1	110	1
Madhya Pradesh	142	14	117	14	90	17	76	18	29.2	11	22.6	12	40.3	13	441	11	407	12	379	12
Maharashtra	79	3	60	4	48	2	36	2	64.3	2	78.4	3	58.8	7	166	5	169	3	149	3
Orissa	135	13	124	15	97	18	75	17	36.1	9	43.7	10	51.8	9	346	10	424	13	358	10
Rajasthan	108	10	79	11	81	15	68	14	21.1	12	17.3	15	26.5	17	508	12	501	14	445	13
Tamil Nadu	91	6	57	3	52	3	37	3	65.1	1	88.8	1	80.9	1	131	2	167	2	134	2
Uttaranchal	nr	nr	nr	nr	52	3	42	5	nr	nr	40.9	11	60.0	6	nr	nr	nr	nr	nr	nr
Uttar Pradesh	150	15	97	13	84	16	73	16	19.8	13	20.2	14	23.0	18	606	15	539	15	517	15
West Bengal	91	6	71	8	52	3	38	4	34.2	10	43.8	9	64.3	4	303	9	218	7	194	7
All India	110		80		70		58		35.4		42.0		43.5		398		327		301	

Note: "nr" = "not relevant" because Chattisgarh, Jharkhand and Uttaranchal were created out of Madhya Pradesh, Bihar and Uttar Pradesh respectively in the year 2000.
[a] For MMR, the data for the newly created states are included in the original larger states.
Sources: National Family Health Surveys (1992/3, 1998/9 and 2005/6); Sample Registration System, Registrar General of India.

Table 16.6. Incidence of communicable diseases, Punjab and other Indian states

State	Tuberculosis (no. of cases per 100,000 population)				Malaria (no. of cases per 100,000 population)				Adult HIV prevalence, 2006 (% of adult population)	
	1998/9		2005/6		1992/3		1998/9			
	Detected	Cured (%)	Detected	Cured (%)	Incidence	Rank	Incidence	Rank	%	Rank
Punjab	207	95	201	100	2,546	8	1,082	12	0.12	12
Andhra Pradesh	592	75	449	91	1,944	9	4,851	3	1.05	1
Assam	710	50	654	93	2,707	7	2,974	9	0.03	18
Bihar	989	84	797	92	1,428	10	3,788	7	0.16	10
Chattisgarh	nr	nr	310	90	nr	nr	nr	nr	0.17	8
Gujarat	438	82	538	98	3,228	6	4,449	4	0.43	4
Haryana	358	88	340	95	933	11	2,093	10	0.10	16
Jharkhand	nr	nr	659	91	nr	nr	nr	nr	0.11	13
Karnataka	269	82	141	96	457	14	600	13	0.81	2
Kerala	526	77	275	97	112	15	56	15	0.13	11
Madhya Pradesh	602	86	353	94	4,728	4	10,015	1	0.11	13
Maharashtra	282	82	321	97	3,742	5	4,098	6	0.74	3
Orissa	833	85	418	89	5,148	2	7,414	2	0.22	7
Rajasthan	397	86	371	97	5,103	3	4,099	5	0.17	8
Tamil Nadu	479	89	508	95	576	13	380	14	0.39	5
Uttaranchal	nr	nr	345	95	nr	nr	nr	nr	0.08	17
Uttar Pradesh	551	77	450	94	7,395	1	3,552	8	0.11	13
West Bengal	492	59	605	95	678	12	1,482	11	0.30	6
All India	544	79	445	94	3,324		3,697		0.36	

Note: "nr" = "not relevant" because Chattisgarh, Jharkhand and Uttaranchal were created out of Madhya Pradesh, Bihar and Uttar Pradesh respectively in the year 2000.
Sources: National Family Health Surveys (1992/3, 1998/9 and 2005/6); National AIDS Control Organization.

the delivery of health services, compared with the national average of 12–14%. This study clearly reveals the quality deficit in health service delivery in Punjab.

As in education, Punjab does not lack in infrastructure facilities for health, but the mechanisms for delivering health care pose a governance challenge. Political interference with routine appointments produces a system in which discipline cannot be enforced, accountability is absent, and employment guarantees rather than service guarantees become the driving force. Moreover, the vertically compartmentalized approach, with separate and parallel programs for immunization, malaria, TB, HIV, etc. leads to frequent mismatches between infrastructure, equipment, supplies, and manpower, as each program seeks to achieve its process targets. Thus, a laboratory technician for the TB control program is unwilling to test for malaria; a radiographer is posted where there is no X-ray machine; no midwives are posted to a primary health center which is meant for institutional delivery.

Since the quality of primary health care rests heavily on health workers who visit the villages and on a system which ensures that they do their job within an integrated program, it is not surprising that the key services, such as immunization and maternal, children's and ante-natal care, have suffered. There is also a need to spread awareness of the importance for better health of sanitation, safe drinking water and institutional/assisted deliveries.

In April 2006, Punjab initiated a system of outsourcing primary health care to local/private care providers in its 1,200 Subsidiary Health Centres/Dispensaries in rural areas under the overall supervision of the *zila parishads* (district councils). It remains to be seen how effectively these doctors can be monitored by the *zila parishads*.

We shall not be able to assess the extent to which the MDG with respect to sanitation and safe drinking water has been attained, because the available data are very poor and the definition of what constitutes "safe" drinking water even poorer.[8] It is worth noting that a United Nations study of 122 countries ranked India third from the bottom on drinking water.

II. Economic Growth

Economic growth has suffered a major setback in Punjab over the past two decades. Punjab had a glorious record of Gross State Domestic Product (GSDP) growth of 5% per annum in the three decades of the 1960s, 1970s and 1980s. By contrast, India's

[8] According to Indian norms, access to an improved water supply exists if at least 40 liters per capita per day of safe drinking water are provided within a distance of 1.6 km.

Table 16.7. Annual growth in GDP by sector, Punjab and all India, 1960/1961 to 2006/2007 (%)

	Agriculture	Industry	Construction	Services	GSDP
PUNJAB					
1960/1 to 1970/1[a]	4.9	6.7	4.5	5.5	5.2
1970/1 to 1980/1	3.2	8.5	0.9	7.1	4.7
1980/1 to 1990/1	4.9	9.1	1.0	4.5	5.3
1990/1 to 2000/1	3.1	6.2	7.9	5.7	4.7
2001/2 to 2006/7	2.0	2.9	9.8	5.8	4.1
2001/2	0.8	−4.2	7.7	5.3	2.0
2002/3	−1.2	6.0	−0.4	6.1	3.1
2003/4	5.8	4.3	8.0	4.7	5.2
2004/5 (provisional)	2.2	3.2	23.8	5.7	4.9
2005/6 (quick estimates)	1.9	4.2	9.8	6.3	4.6
2006/7 (advance estimates)	2.4	3.9	9.8	6.6	4.9
All INDIA					
1960/1 to 1970/1	2.5	5.4	5.6	4.8	3.8
1970/1 to 1980/1	1.8	4.4	3.3	4.4	3.2
1980/1 to 1990/1	3.5	6.7	4.7	6.6	5.4
1990/1 to 2000/1	2.8	5.7	5.1	7.3	5.6
2001/2 to 2006/7	3.0	7.1	10.5	8.9	7.3
2001/2	6.3	2.4	4.0	7.2	5.8
2002/3	−7.2	6.8	7.9	7.4	3.8
2003/4	10.0	6.0	12.0	8.5	8.5
2004/5	0.0	8.4	14.1	9.6	7.5
2005/6	6.0	8.0	14.2	9.8	9.0
2006/7 (revised estimates)	2.7	11.0	10.7	11.0	9.4

[a] Growth of NSDP.
Source: National Accounts Statistics and Statistical Abstract of Punjab.

GDP grew at a slow rate of $3\frac{1}{2}$% per annum in the first two decades, and caught up with Punjab's growth rate only in the 1980s. Beginning with the 1990s, economic growth in Punjab has fallen below the growth rate of the Indian economy, and the divergence has widened with time (Table 16.7 and Figure 16.4).

A notable aspect of Punjab's economic development since 1960/1 has been the very slow structural change generated by the process of industrialization in the state. The underlying driver of economic transformation was agriculture, and the development of a sound industrial strategy did not find favor with the political economy of the state. Development of value-added products using the rich agricultural base was the natural course to follow, but except for dairy and some milk-based products, this potential was not tapped. This meant that as agricultural productivity increased, there were not enough employment opportunities for labor released from the farms. As youth unemployment rose, only some could find outlets

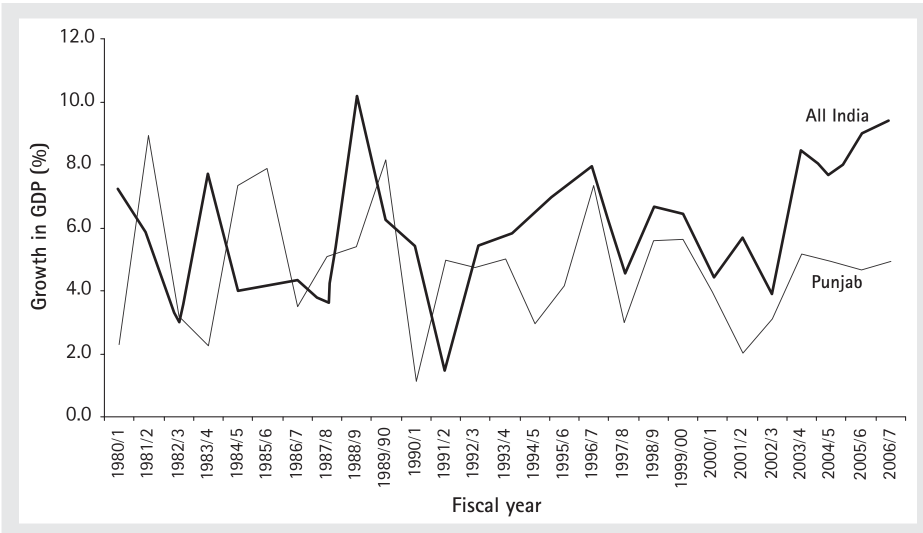


Fig. 16.4. Annual growth in GDP, Punjab and all India

through migration to industrialized countries. The industrial deficit can be clearly seen in Figure 16.5, which plots the industrial share of GSDP against per capita income across the major states of India.

With its excellent irrigation facilities, Punjab was ideally placed in the late 1960s to usher in the Green Revolution. As government policy tried to plug market

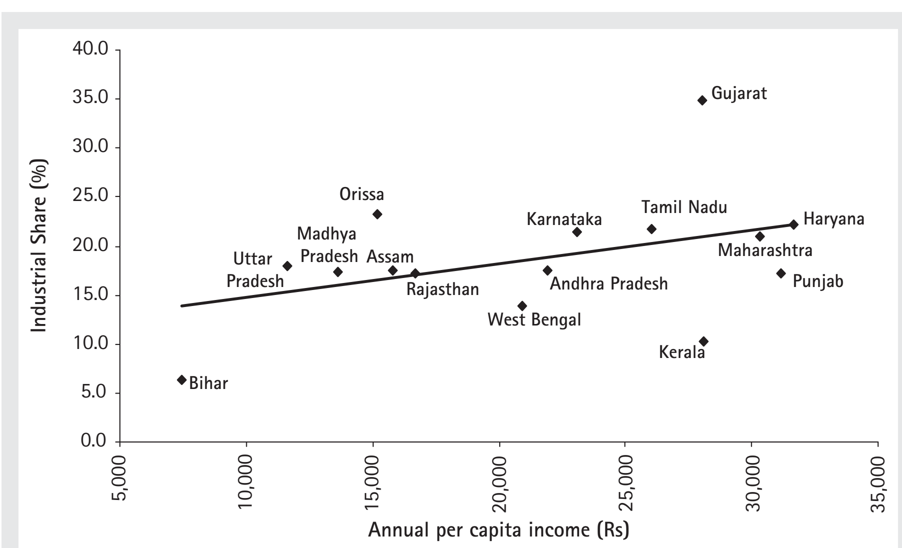


Fig. 16.5. Share of industry in GSDP, major states of India, 2004/2005

failures by encouraging the use of high-yielding varieties of imported wheat, promoting fertilizers, and spreading extension services, there was substantial alignment between the case for publicly implementing the technology of the Green Revolution and the political centrality of the middle peasantry in Punjab. Notwithstanding the agrarian focus of the policy regime and the constraints that arise from being a landlocked state, the Punjabi entrepreneurial spirit, along with the high-spending propensity of Punjabis, drove the process of industrialization, achieving an industrial growth rate of 8.5% per annum in the 1970s. It reflected the activities of mostly small-scale units in metal-based light engineering industry, hand tools, hosiery and knitwear, leather goods, sports goods, and auto components.

The growth process in Punjab faced a number of challenges in the 1980s. The worsening economic conditions in the Soviet Union had an adverse impact on the woolen hosiery industry, which had benefited from the Indian government's Special Trade Arrangements with the Soviet Union. The phasing-out of the freight equalization pool policy threatened the economic viability of the light engineering units.[9] Rising militancy in the political field vitiated the investment climate more generally. But the continuing momentum of agricultural growth, at 4.9% per annum, was a major factor boosting industrial growth in Punjab to 9.1% per annum in the 1980s. The overall growth in GSDP was 5.3% per annum. The rapid rise in farm incomes in Punjab actually created a situation where not only demand for labor on the farms declined, but also children of rich farmers were disinclined to do the work that still had to be done manually on the farms. Farm labor was increasingly supplied by migrant workers from Bihar and Orissa.

During the 1980s, GDP growth in the Indian economy accelerated to 5.4% per annum (a little higher than the 5.3% per annum of Punjab), responding to the introduction of market-oriented policy reforms. When the government of India was faced with the balance of payments crisis of 1990/1, it set in motion a series of wide-ranging policy reforms, emboldened by its experience of the 1980s, and reaped the benefits by sustaining the growth momentum of the earlier decade.

Punjab, by contrast, was losing ground in the 1990s. Agriculture, its principal engine of growth, was running out of steam as land was becoming more and more scarce and yield per hectare for wheat and rice was stagnating, reflecting technology fatigue (Figure 16.6). A major deceleration in agricultural growth after the 1980s resulted in a slowdown in the industrial sector. Moreover, unlike many other states, who started gearing up their administrative machinery to attract private investment in industry, Punjab failed to take advantage of the new industrial opportunities that were opening up. Industrial growth in Punjab declined sharply from 9.1% per annum in the 1980s to 6.2% per annum in the 1990s and further to 2.9% per annum in the first five years of the new millennium. Some of the existing industrial

[9] The freight pool equalization policy of the government of India in the pricing of steel had helped offset the locational disadvantage of the small-scale units in the light engineering sector in Punjab.

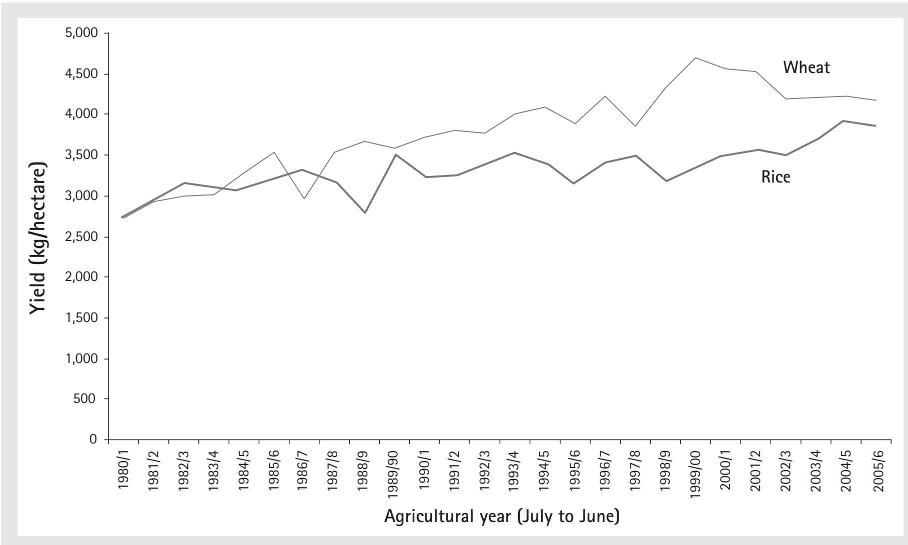


Fig. 16.6. Yield per hectare, wheat and rice, Punjab

firms chose other states as locations for expansion. New investments, including FDI in the automotive sector and auto parts, went to the states in the South and the West, which encouraged vendor development and helped create a modern auto components sector with the emphasis on quality. Punjab missed out on this transition.

A measure of the industrial opportunities lost in Punjab in the 1990s in textiles was the progress made by the cluster of hosiery and textile units at Tirupur in Tamil Nadu, which started a good 20 years later than Ludhiana but made tremendous strides in the 1990s towards global competitiveness. During 1996 to 1998, exports from Tirupur grew at 12% per annum, while exports from Ludhiana grew at only about 3% per annum (Rathore *et al.* 1999). In recent years cotton has surpassed woolen hosiery in Ludhiana and these units are also operating in a larger, year-round market, with the emphasis on high-value products.

More recently, the investment climate in Punjab has suffered a major setback as a result of the decision of the central government in 2003 to offer a number of attractive fiscal concessions to industrial investments in the states of Himachal Pradesh, Uttaranchal and Jammu and Kashmir—all neighboring states of Punjab. This has not only had an adverse impact on new investment in industry in Punjab, but has also led to investment flight, particularly in footloose industries.

On the other hand, Punjab has had tremendous success in dairy farming. Nestlé's development of the Moga district through building supply chains working with producers is a shining example of what the private sector can do for rural development.

Punjab has done better in services than in industry, but has not fully shared in the resurgence of the growth in services at the all-India level. Considering that its locational handicap is much less relevant for the IT sector and that good-quality telecommunications infrastructure and high-quality education are the two basic requirements, the state is now gearing up to exploit these opportunities in Mohali near Chandigarh.

Overall, Punjab shows a consistent shortfall in growth *vis-à-vis* the Indian economy during the reform period beginning with the early 1990s. In the four years from 2003/4 to 2006/7, the shortfall widened so much that the average annual growth rate of 4.9% in Punjab was nowhere near the GSDP average annual growth rate of 8.6% for the Indian economy as a whole (see Figure 16.4). What is more, the sustainability of even this slower growth rate is becoming a serious problem in view of the growing environmental deficit in Punjab.

III. Challenges Ahead

Punjab faces enormous challenges in the years ahead in sustaining the high growth rates of the past, let alone in striving for higher growth. The most significant challenge comes from the severe deterioration in the natural resource base over the years.

The absence of economic pricing of water in Punjab has led to overuse of water and increasing cultivation of paddy, which is highly water-intensive. The policy of free electricity for farmers has led to excessive use of pump sets for accessing groundwater and there is little incentive to conserve either water or electricity.[10] As a result, Punjab has the highest percentage of groundwater exploitation in the country, with over 85% of its 141 blocks now classified as "overexploited" or dark blocks. As the water table declines, the economic cost of pumping water also rises, along with its environmental cost.

Punjab, along with other states, suffers from the central government's irrational system of fertilizer subsidy, where urea is over-subsidized. This has created major problems of imbalance in the use of fertilizers, with inadequate application of potassium and phosphate fertilizers and the consequent effect on soil quality and yield. Excess use of nitrogenous fertilizers has also led to depletion of micronutrients in the soil. Even though the need for replenishing micronutrients is well documented scientifically, the fertilizer subsidy regime, in which the subsidy is limited to major fertilizers only, creates the wrong incentives.

[10] It is estimated that 45% of the electricity demand in the agricultural sector is concentrated in the paddy season.

With the acceleration of growth and the associated rise in incomes in India, consumption patterns are predictably moving away from food grains to vegetables, fruits, pulses, milk, eggs, etc. While institutional mechanisms need to be put in place to mitigate the risk of high-value crops, the challenge of stagnation of yield in wheat and rice (Figure 16.6) also calls for research directed at better and more robust seeds for high-value crops. Investments in cold chains will also help improve returns and reduce risk for farmers. The National Horticultural Mission provides an opportunity to farmers in Punjab to tap resources from the government of India to move towards horticulture farming.

Modernization of retail services offers a new set of spectacular opportunities for diversifying into high-value crops. Growing interest by private parties such as Reliance, Bharati and ITC in marketing, logistics and retail indicates the potential of this sector. Avenues for effective marketing of crops are crucial if high-value agriculture is to find its place in Punjab, but there is stiff resistance from Arthias (the established intermediaries). After one failed attempt, the government of Punjab is again working on an amendment to the Agriculture Produce Marketing Committees Act to provide wider scope to the private sector in the marketing of crops. A number of other states have already amended this Act, in line with the "model amendment" suggested by the government of India.

Emphasis on agro-based industries, e.g. food processing and cotton textiles, will provide backward links from industry to agriculture. A cluster approach through provision of infrastructure will also help realize the potential of the existing small-scale units in cotton textiles, auto components, hand tools and leather products.

The success of industrial ventures is crucially dependent on the availability of good-quality infrastructure, particularly power and transport. The correct pricing of power assumes great significance in this context. Punjab also needs to maintain its good network of roads and bridges and supplement this with modern airports, such as the new international airport at Amritsar. A significant development for Punjab has been the government of India's decision to extend the proposed dedicated rail freight corridor from Mumbai, Kolkatta and Delhi up to Ludhiana. This will considerably reduce the inland haulage cost of export cargo from Punjab.

A strategy to revive economic growth in Punjab will therefore have to face the dual challenge of bridging the environmental deficit and the industrial deficit, while building synergies between the growth of the industrial sector and the diversified agricultural sector. The consistently high fiscal deficits also have a potentially adverse impact on growth, quite apart from limiting the capacity of the state to increase its spending on health and education (Figure 16.7). Finally, the serious issue of governance will have to be resolved through reforms in policies and institutions which have a bearing on the functioning of the system of public service delivery.

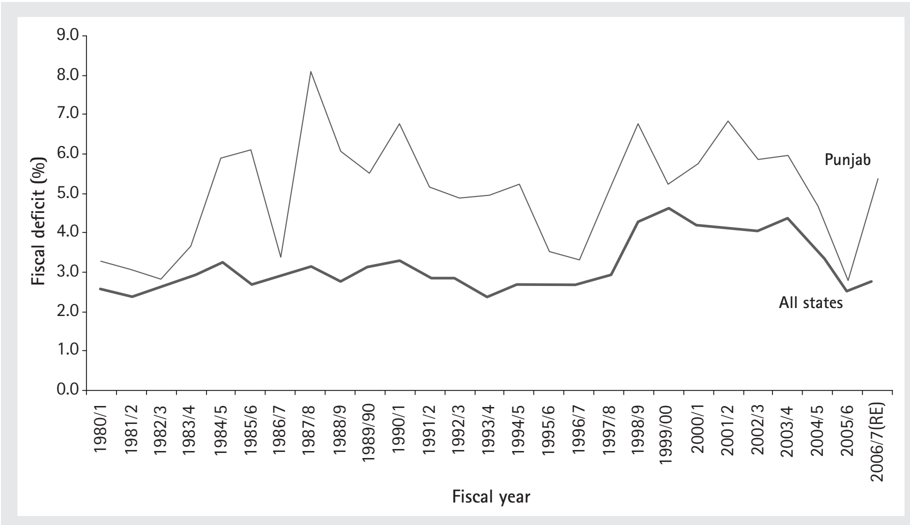


Fig. 16.7. Fiscal deficits, Punjab and all states

References

Chaudhury, Nazmul, Hammer, Jeffrey, Kremer, Michael, Muralidharan, Karthik, Rogers, Halsey F. (2006), "Missing in Action: Teacher and Health Worker Absence in Developing Countries", *Journal of Economic Perspectives*, 20: 91–116.

Deaton, Angus, and Drèze, Jean (2002), "Poverty and Inequality in India A Re-Examination", *Economic and Political Weekly*, 7 September: 3729–48.

HRD (Ministry of Human Resource Development) (2007), *Elementary Education in India: Progress Towards UEE* (New Delhi: HRD).

Kremer, Michael, Chaudhury, Nazmul, Rogers, Halsey F., Muralidharan, Karthik, Hammer, Jeffrey (2005), "Teacher Absence in India: A Snapshot", *Journal of the European Economic Association*, 3: 658–67.

Mehta, Arun C. (2005), *Elementary Education in Unrecognized Schools in India: A Study of Punjab based on DISE 2005 Data* (New Delhi: NIEPA).

—— (2005/6), *Elementary Education in India: Where Do We Stand?*, State Report Cards 2004 (New Delhi: NIEPA and Government of India).

Muralidharan, Karthik, and Kremer, Michael (forthcoming 2008), "Public and Private Schools in Rural India", in *School Choice International* ed. Paul Peterson and Rajashri Chakrabarti (Cambridge, Mass.: MIT Press).

National AIDS Control Organization (2006), "Note on HIV Sentinel Surveillance and HIV Estimation", available at <www.nacoonline.org/Quick_Links/Publication/>.

NFHS (various years), *National Family Health Survey* (Mumbai: International Institute for Population Sciences).

NCERT (various years), *Learning Achievements of Students* (New Delhi: NCERT).

Paul, Samuel, *et al.* (2004), "State of India's Public Services: Benchmarks for the States", *Economic and Political Weekly*, 28 February: 920–33.

Pratham (various years), *Annual Status of Education Report* (ASER), available at <www.pratham.org/aserrep.php>.

Rao, Govinda, and Choudhury, Mita (2008), *Inter-State Equalisation of Health Expenditures in Indian Union* (New Delhi: National Institute of Public Finance and Policy).

Rathore, Sharma, and Saini (1999), *Comparative Study on Entrepreneurship in Hosiery Industry in Ludhiana (Punjab) & Tirupur (Tamil Nadu)* (Chandigarh: National Institute of Technical Teachers' Training and Research).

SRS (2006), *Sample Registration System* (New Delhi: Office of the Registrar General and Census Commissioner).

Sankar, Deepa, and Kathuria, Vinish (2004), "Health System Performance in Rural India, Efficiency Estimates across States", *Economic and Political Weekly*, 27 March: 1427–33.

Shantayanan, Devarajan, and Shah, Shekhar (2004), "Making Services Work for India's Poor", *Economic and Political Weekly*, 28 February: 907–19.

Sinha, Amarjeet (2005), *India: Democracy and Well-Being, Meeting People's Health Needs* (New Delhi: Rupa and Co.).

UNDP (2004), *Human Development Report: Punjab (2004)* (New York: UNDP).

CHAPTER 17

GROWTH, DISTRIBUTION AND INCLUSIVENESS

REFLECTIONS ON INDIA'S EXPERIENCE

MONTEK S. AHLUWALIA

Distributional issues have long been central to development policy but the intensity of debate has increased in recent years, fed by growing suspicion that the pursuit of market-oriented policies in a globalized world benefits only the few at the expense of the many. The debate is especially vigorous in India, as befits "argumentative Indians", to use Amartya Sen's evocative phrase. After a long period of low to modest growth, India has made an impressive transition to higher growth, but performance on the distributional front has been less impressive. Distributional demands are multiplying, reflecting higher expectations generated by greater awareness and empowerment, and also by the success of the growth strategy, which has produced many visible gainers. Responding to these demands is a major challenge. Failure on the distributional front will undermine support for the policies which have stimulated growth, and yet, rapid growth is certainly necessary, even if not sufficient, to ensure that distributional objectives are met.

I. Evolving Perspectives on Growth and Distribution

The relationship between economic growth and distributional outcomes in India has been perceived differently over time. A brief review of this evolution of perceptions is of interest if only because that evolution foreshadowed changing perceptions in the development community internationally.

I.1 Growth as an Instrument for Raising Living Standards

The early perception of the relationship between growth and the standard of living, implicit in the development plans of most developing countries in the 1960s, was that growth is necessary and also sufficient for an improvement in living standard of the general population. In India's case, this approach can be traced back to the National Planning Committee set up by the Indian National Congress in 1938 under the chairmanship of Pandit Jawaharlal Nehru. The Committee observed that the average income was so low that the national income must be increased two- to three-fold in ten years to ensure an irreducible minimum standard for everyone.[1] This implied a target growth rate of between 7.2 and 11.7% per year and implicitly assumed that distribution would not deteriorate and offset the benefits of growth.

India's early Five-Year Plans were basically growth-driven, though the targets were more modest, at around 5% per year. However, doubts about whether the benefits of growth were reaching the mass of the people began to surface by the early 1960s. Prime Minister Nehru, introducing the Third Five-Year Plan (1961/2 to 1965/6) for discussion in Parliament on 22 August 1960, observed that per capita income had increased by 20% over the decade of the first two Five-Year Plans but noted that "a legitimate query is made—where has this gone?" He went on to observe that when he addressed gatherings in the villages he could see that the people were better fed and clothed but "that does not apply to everybody in India. Some people have hardly benefited. Some people may even be facing various difficulties" (quoted in Planning Commission 1964).

The Planning Commission responded to these concerns by producing a detailed strategy for bringing everyone above a minimum level of living within 15 years, fixing the minimum at Rs20 per month, excluding expenditure on health and education, which were expected to be provided free by the state (see Planning Commission 1964). The strategy recognized that there were limits to what could be achieved by redistribution in a poor country, and growth was therefore essential for a sustained increase in living standards. However, it also recognized that

[1] The Committee never submitted a report because its work was overtaken by the outbreak of the Second World War. However, the deliberations of the Committee are reflected in Shah (1949).

distribution might worsen and the incomes of the lower deciles would therefore rise more slowly than the total. The share of the third decile from the bottom was projected to decline from 4.5% of total income in 1961 to around 4% by 1976, and on this basis the GDP needed in 1976 to ensure that all individuals in the third decile came up to the minimum level required GDP to grow at 7% per year over the 15-year period from 1961 to 1976. This was judged to be feasible. Individuals in the two lowest deciles would still remain below the minimum level on this projection and bringing this group to minimum levels would require suitable income transfers. This scale of redistribution was felt to be achievable without disrupting the momentum of growth.

This strategy represented a significant departure from the past in several respects. First, it defined the distributional objective not in terms of equity as such, but in terms of elimination of absolute poverty. This approach was adopted in the international development community ten years later when the World Bank began to focus on absolute poverty. Second, it defined the minimum not simply in terms of access to private consumption goods but also in terms of access to health and education, elements that came to occupy center stage much later. Finally, it explicitly recognized that income distribution could deteriorate. It did not, however, link the growth strategy to distributional outcomes, treating the latter as an exogenous development, which implied that faster growth would bring more people above the poverty line.

I.2 Emphasizing Anti-Poverty Programs

As it turned out, the 7% growth target proved unattainable. The Third Plan (1961/2 to 1964/5), which was approved before the fifteen year strategy was formulated, had set a lower target of 5.6% and actually achieved only 2.8%. This poor performance was attributed at the time to external shocks, including the war with China in 1962, followed shortly thereafter by the war with Pakistan in 1965, which was followed by two droughts. The more basic question of whether the fault might lie with the economic policies being followed was not addressed at the time.

Inevitably, failure on the growth front generated pressure for an alternative strategy focusing on programs directly aimed at helping the poor. Political developments in the country also encouraged this shift. The Congress Party split in 1969 following an attempt by the old guard to rally the right wing against Prime Minister Indira Gandhi. A year later the Prime Minister dissolved Parliament to seek a fresh mandate. The general election in 1971 was fought and handsomely won on the politically attractive slogan *Garibi hatao* (Remove poverty).[2]

[2] The slogan was also an effort to shift the political debate from personalities to issues. As Mrs Gandhi put it "woh kehte hain Indira hatao, mein kehti hun Garibi hatao" (They say remove Indira, I say remove poverty). See Guha (2007).

A variety of anti-poverty programs was launched in the 1970s, aimed at providing employment and other types of income support to target groups such as rural landless labor, small farmers and self-employed individuals in the informal sector. A Minimum Needs Programme was also launched in 1972/3, aimed at the provision of education, rural health, rural roads, a rural water supply and rural housing.

The focus on anti-poverty programs foreshadowed a similar shift in the international development community a few years later when the World Bank began to emphasize the need for targeted anti-poverty programs as part of an overall development strategy. The advocacy of "Basic Needs" by Mahbub ul Haq in the World Bank in the second half of the 1970s, and later in the UNDP, is also reminiscent of the Indian focus on Minimum Needs Programs in the early 1970s. However, India's anti-poverty strategy was implemented in an environment in which growth rates remained low. The Fifth Five-Year Plan (1974/5 to 1978/9) was formulated in an environment of severe shortage of foreign exchange following the first oil crisis and the drying up of international assistance, and this led to a scaling down of the growth target to 4.7%. An alternative approach to policy, liberalizing controls on domestic investment and foreign trade so as to achieve higher levels of efficiency which might have improved growth performance, had been advocated by Bhagwati and Desai (1970), but these ideas were not accepted until much later. Instead, the Plan accepted lower growth targets and sought to pursue distributional objectives through a sharper redistribution of consumption in favor of the poor, though the instruments for achieving this redistribution were not spelled out.[3]

As it turned out, even the reduced growth target could not be met. As shown in Table 17.1, the average growth rate in the two decades from 1960 to 1980 was only 3.5%, famously called the "Hindu rate of growth" by Raj Krishna. The anti-poverty programs were plagued by leakages and weak implementation and did little to counter weak growth. The percentage of the population in poverty actually increased marginally from 53% in 1970/1 to 55% in 1973/4 (see Table 17.2) and then declined to 51% in 1977/8, which was only marginally lower than in 1970/1.

I.3 Reforms to Accelerate Growth

The realization that India's growth performance was much poorer than that of other developing countries in East and South-East Asia led to a serious reconsideration of policy in the 1980s. Several expert committees recommended liberalization of various aspects of domestic controls, industry associations lobbied actively for such

[3] See Tendulkar (1974). The lower growth was not the consequence of a trade-off. Rather, the sharper redistribution was imposed as a response to the lower growth target without considering whether it might itself reduce growth.

Table 17.1. Growth and other macroeconomic indicators, India, 1960–2006

	Growth in GDP (%)	Gross domestic investment (% of GDP)	Gross domestic savings (% of GDP)	Export of goods and services (% of GDP)
1960–80 (annual average)	3.5	17	16.5	4.6
1980–90 (annual average)	5.6	21.4	19.4	6.1
Eighth Plan targets (1992–6; annual average)	6.5	24.4	23.1	10.3
Ninth Plan targets (1997–2001; annual average)	5.5	24.2	23.6	12.1
Tenth Plan targets (2002–6; annual average)	7.8	31.2	31.3	18.4
2002/3	3.8	25.2	26.4	14.7
2003/4	8.5	28.0	29.7	15.4
2004/5	7.5	31.5	31.1	18.4
2005/6	9.6	35.9	34.8	20.7
2006/7	9.4	35.1	34.7	22.8

Sources: GDP, savings and investment calculated from Government of India, *Economic Survey 2008* and earlier issues; exports from Database on Indian Economy, Reserve Bank of India.

Table 17.2. Trends in poverty, India, 1970/1971–2004/2005

	Population in poverty (%)			Numbers of poor (millions)		
	Rural	Urban	Total	Rural	Urban	Total
Estimates based on URP[a]						
1970/1[b]	54.8	45.0	52.9	not available		
1973/4	56.4	49.0	54.9	261.3	60.0	321.3
1977/8	53.1	45.2	51.3	264.3	64.6	328.9
1983	45.7	40.8	44.5	252.0	70.9	322.9
1987/8	39.1	38.2	38.9	231.9	75.2	307.1
1993/4	37.3	32.4	36.0	244.0	76.3	320.3
2004/5	28.3	25.7	27.5	220.9	80.8	301.7
Estimates based on MRP[c]						
1999/2000 (MRP modified)	27.1	23.6	26.1	193.2	67.0	260.2
2004/5 (MRP)	21.8	21.7	21.8	170.3	68.2	238.5

[a] URP = the Uniform Recall Period, in which the consumer expenditure data for all items are collected from a 30-day recall period.
[b] The estimate for 1970/1 is taken from Ozler *et al.* (1996) and is not strictly comparable with the other numbers, which are from Planning Commission documents. The comparability of the 1970/1 figures can be judged from the fact that the estimate of total poverty for 1973/4 from Ozler *et al.* is 54.1%, i.e. it also shows an increase between 1970/1 and 1973/4 but less than indicated by the official estimate of 54.9%.
[c] MRP = Mixed Recall Period, in which the consumer expenditure data for five non-food items, namely clothing, footwear, durable goods, education and institutional medical expenses, are collected from a 365-day recall period and the consumption data for the remaining items are collected from a 30-day recall period.
Source: Planning Commission, Government of India.

changes. There was also an accumulation of academic work pointing in the same direction.[4] A number of steps were taken in that decade to liberalize domestic controls on industry, encourage technological modernization, and begin a process of rationalizing the tax system. Though an important break from the past, these initiatives were non-systemic in the sense that the basic system of multiple controls over industrial investments, foreign trade and foreign direct investment remained in place but they were administered more liberally, giving greater room for private sector expansion and modernization.

Poverty alleviation continued to be an important objective in the 1980s and the anti-poverty programs adopted in the 1970s were expanded. The strategy therefore continued to be "growth plus", but with a focused effort at stimulating the growth component, which was visibly missing. The weaknesses of the anti-poverty programs were also frankly acknowledged. Prime Minister Rajiv Gandhi famously commented that of every rupee spent on the poor, he doubted whether more than 15 *paisa* reached the intended beneficiaries. The economy responded well to the new initiatives and growth accelerated to 5.6% in the 1980s. As shown in Table 17.2, poverty also declined in this period.

The next stage in the evolution of economic policy came in 1991 when Dr Manmohan Singh, as the Finance Minister of the newly elected Congress government, introduced a more comprehensive program of economic reforms. These reforms were "systemic" in the sense that they explicitly recognized the need to change the system of controls to give a much larger role to the market and the private sector and also to integrate India with the world economy. Unlike the 1980s, fiscal discipline was emphasized in recognition of the fact that the 1991 crisis was seen to be the result of fiscal expansionism in earlier years. There was also a conscious push for external liberalization, involving the lowering of trade barriers and encouraging foreign investment. Though systemic in conception, the reforms were implemented cautiously in a series of measured steps, designed to limit the disruptive impact at any given point.

Precisely because of their systemic nature, the reforms of the 1990s were much more controversial. This was also the period when there was an active international debate on the impact of "Washington Consensus"-type reforms in Asia and Latin America. Criticisms of the reforms in India echoed the international debate and focused on (1) accusations of the imposition of World Bank–IMF conditionality, (2) worries that a flood of imports would lead to a collapse of domestic industry, (3) concerns over the potential negative impacts of trade liberalization and fiscal discipline on poverty and employment, and (4) the risks of environmental degradation and, more recently, the displacement of tribal communities as a

[4] Bhagwati and Desai (1970) had done a detailed analysis of the inefficiencies and costs of the control mechanism. Ahluwalia (1986) showed that the low growth rates in industry were due not to low rates of investment but to declining total-factor productivity growth, which in turn was linked to the system of government controls.

result of developmental projects, since their property rights are often not legally defined.

Inevitably, the debate degenerated at times into sterile confrontation between pro-market and anti-market positions, but the issues underlying the debate were genuine and needed to be addressed. Whether the economic reforms would indeed accelerate growth or lead to a "lost development decade" as in Latin America was clearly relevant. Whether they would have adverse distributional consequences was also a valid question. Reforms are expected to increase productivity by bringing about structural shifts across sectors, firms and technologies, and it is easy to imagine differing impacts which can adversely affect some groups, at least in the short run. Adverse impacts on some groups may well be offset by positive effects on others but the blame for negative effects is invariably laid at the door of the government, while the credit for positive effects is not so easily claimed.

I.4 Evolution of Distributional Objectives

Distributional objectives have also evolved over time becoming increasingly multidimensional. From the very beginning, planners acknowledged the need to achieve a minimum level of consumption, with health and education being provided free for the bulk of the population, but in practice the focus was primarily on providing income support and subsidized food to the poor in order to bring them up to minimum levels of consumption expenditure.[5] As the economic reforms led to a withdrawal of the state from areas better left to market forces, the core function of government came to be identified with provision of health and education services and the deficiencies in these areas received increasing attention. In time, the set of publicly supplied services crucial for distributional objectives was expanded to include the rural road network, rural electrification, rural drinking water and sanitation, with targets for universal coverage for delivering these services.

More recently, distributional objectives have been defined to reach beyond ensuring adequate incomes and providing essential services for the poor, to include the much more challenging task of giving excluded groups a true sense of inclusion and empowerment. Excluded groups in this context are not defined in terms of income or consumption levels alone, but in terms of scheduled castes, scheduled tribes, religious minorities, etc. that suffer from societal discrimination. These communities would be heavily represented in any purely income-based classification of exclusion, but the evolution of a distinct group identity—itself the consequence of

[5] The Indian Constitution, adopted in 1950, called for universal elementary education to be achieved within ten years. Planning Commission (1962) had set a target of having every child in school by 1976, an objective that is not yet fully accomplished 30 years later.

societal discrimination—leads naturally to the demand for group empowerment and participation in decision-making. Poverty alleviation, strictly defined, would focus only on bringing individuals above some poverty line, but what is demanded now is a fair sharing of opportunities at all levels. It is worth noting that the demand for a fair share of opportunities is actually quite distinct from the demand for greater equality, since it can be achieved without any reduction in inequality, requiring only that the relevant group making the demand is proportionately represented at all levels of income.

Finally, the traditional demands for a reduction in inequality of incomes and income-earning opportunities have also resurfaced as globalization and the electronic media have made the public much more aware of both the outstanding successes and the abject failures. News that India has the fourth largest number of billionaires in the *Forbes Magazine* list, after the US, Germany and Japan, competes for attention with news of crop failures and severe distress in drought-affected areas and non-delivery of essential public services in many parts of the country. The fact that there is a historically unprecedented broad-based improvement in the living standards of hundreds of millions of people receives less attention than it deserves, because of the perception that inequality has increased.

The next two sections evaluate how India has fared with respect to economic growth and distributional objectives in the post-reform period.

II. The Experience of Growth

Economic reforms have clearly been good for growth, though this was not so clearly evident until only a few years ago. The reforms of the 1980s were applauded because, after two decades of slow growth at 3.5% per year, the economy grew at 5.6% per year in the 1980s (see Table 17.1). Growth accelerated further to 6.5% in the period 1992–6, contradicting fears that the economic reforms would have a negative effect on growth. However, growth faltered in the second half of the decade and for the 1990s as a whole the economy averaged only 5.8% growth, just marginally higher than in the 1980s.

Rodrik and Subramanian (2004a) have attributed the absence of acceleration in the 1990s to the fact that the reforms of the 1980s, focusing on limited domestic liberalization with a more supportive attitude towards the large domestic private sector corporations, were the really important reforms and the subsequent focus on external liberalization in the 1990s, though much favored by international organizations, was much less important. This argument can be faulted on several grounds.

First, the reforms of the 1980s, though undoubtedly important, did not put the country on a sustainable growth path. There was unsustainable fiscal expansion, which led to the balance of payments crisis in 1990. The broader reforms of the 1990s helped remove these weaknesses, producing a much stronger export performance and an expansion in foreign investment inflows to 3% of GDP in 2006/7, both contributing greatly to external sustainability.

Slow growth in the second half of the 1990s was also partly a cyclical phenomenon. Private corporate investment decelerated after an initial exuberant expansion following liberalization. The investment climate was also dampened by the crises in East Asia in 1997 and subsequently in Russia and Brazil, followed by the collapse of the technology boom in the US in 2000. Agriculture also slowed down after 1996, partly reflecting a succession of poor monsoons, but partly reflecting also a failure to address emerging constraints, a failure that is now being corrected (see Planning Commission 2008).

Finally, the modest growth in the 1990s was also partly due to the cautious pace at which reforms were introduced. This reduced the risk of disruptive impacts and also gave time to develop a political consensus in favor of change, but it also had the disadvantage that the economic agents took longer to respond, delaying the full benefit of the reforms. Nevertheless, as pointed out in Ahluwalia (2002), the cumulative effect of the reforms was considerable and this is reflected in the more recent performance. With GDP growth in 2007/8 at 9.0%, India's average growth rate for the last five years is 8.8%, which is the highest for any five-year period.

The foregoing may overstate the underlying trend, since this was a period of cyclical upswing in the global economy which is now being reversed. However, as pointed out in Planning Commission (2008), there is reason to believe that the underlying fundamentals, including especially savings and investment rates, the capacity of the domestic private sector, and also perceptions of India by foreign investors as an investment destination, all show a distinct improvement. Because of this shift, the Eleventh Plan has targeted a growth rate of 9% for the period 2007/8 to 2011/12 and most observers believe this is feasible, provided some critical constraints are effectively addressed.[6]

An important weakness in the recent growth performance is the performance of the agricultural sector. After averaging 3.6% per year between 1980 and 1996, agricultural growth slowed down thereafter to about 2%. This deceleration has undoubtedly had adverse distributional effects and this poor performance

[6] This proviso is important, since it relates to actions that will unlock productivity potential. Rodrik and Subramaniam (2004b) estimate India's underlying growth potential at 7% per annum, with significant upsides related to fuller exploitation of potential. Planning Commission (2008) also asserts that 9% annual growth cannot be taken as a foregone conclusion, and outlines a comprehensive policy agenda to achieve it.

needs to be addressed in the years ahead if distributional outcomes are to be improved.

III. The Experience with Distributional Outcomes

While the growth performance in the post-reform period has been commendable, the public perception of performance on the distributional front is very different. Since distributional objectives take various forms, e.g. concern about the extent of poverty, access to essential services, income inequality, urban–rural balance, etc., it is useful to evaluate performance in each dimension separately before making a composite assessment.

III.1 Poverty Reduction

The most commonly used measure of distributional outcomes is the percentage of the population in poverty, and the available data show that this percentage has declined continuously, both in the period before the reforms and after. However, the decline is modest at best (see Table 17.2) and the percentage in poverty remains high at 27.5% in 2004/5. While the decline is important, the high percentage of the population in poverty is unacceptable, especially considering that the objective in the 1960s was to take everyone above the minimum level by 1976.

The decline in poverty is evident in all states, even the poorest, although the extent of the decline varies (see Table 17.3). Bihar, the poorest state, shows a large decline, from 55% in 1993/4 to 41.1% in 2004/5. Uttar Pradesh, the largest and one of the poorest states, also experienced a substantial decline. However, Madhya Pradesh and Orissa, also poor states, have seen only a modest decline. This has meant that the absolute number of the poor, which has declined modestly in the country as a whole, has actually increased in Madhya Pradesh, Maharashtra, Rajasthan, Orissa and Uttar Pradesh. The concentration of the poor in the five states of Bihar, Madhya Pradesh, Maharashtra, Uttar Pradesh and Orissa increased from 58% of the total population in poverty in 1993/4 to 64% in 2004/5.

An important criticism of India's performance in poverty reduction is that faster growth in the post-reform period does not seem to have produced faster reduction in poverty. In the period 1973/4 to 1983, when per capita net national product (NNP) grew at only about 1.9% per year, poverty declined by more than 1 percentage

Table 17.3. Poverty in the Indian states, 1993/1994 and 2004/2005

State	Population in poverty (%)[a]		Number of poor (million)	
	1993/4	2004/5	1993/4	2004/5
Andhra Pradesh	22.2	15.8	15.40	12.61
Assam	40.9	19.7	9.64	5.58
Bihar[b]	55.0	41.1	49.34	48.55
Gujarat	24.2	16.8	10.52	9.07
Haryana	25.1	14.0	4.39	3.21
Himachal Pradesh	28.4	10.0	1.59	0.64
Jammu & Kashmir	25.2	5.4	2.09	0.59
Karnataka	33.2	25.0	15.65	13.89
Kerala	25.4	15.0	7.64	4.96
Madhya Pradesh[c]	42.5	39.0	29.85	34.07
Maharashtra	36.9	30.7	30.52	31.74
Orissa	48.6	46.4	16.06	17.85
Punjab	11.8	8.4	2.51	2.16
Rajasthan	27.4	22.1	12.85	13.49
Tamil Nadu	35.0	22.5	20.21	14.56
Uttar Pradesh[d]	40.9	33.1	60.45	62.60
West Bengal	35.7	24.7	25.46	20.84
All India	36.0	27.5	320.37	301.72

[a] Poverty ratios are based on URP-consumption distribution.
[b] Includes Jharkhand.
[c] Includes Chattisgarh.
[d] Includes Uttarakhand.
Sources: Government of India Press Information Bureau, "Estimates of Poverty", 11 March 1997, and "Poverty Estimates for 2004–05", 21 March 2007.

point per year.[7] In the next ten years, from 1983 to 1993/4, the growth rate of per capita NNP increased to about 2.8% per year, but poverty declined more slowly at 0.8% per year. In the next ten years, 1993/4 to 2004/5, growth of per capita NNP accelerated further to around 4.4% per year, but the rate of decline in poverty remained the same.

This apparent weakening of the link between growth and poverty reduction has raised questions about whether the reforms have produced a growth process which is inherently less supportive of poverty alleviation than its predecessors. One reason why this may indeed be so is that, as pointed out earlier, agricultural growth slowed down after 1996. The fact that industrial growth in India has not been high enough and also not sufficiently labor-absorbing is another reason. Bhalla (2004) has also argued that there may be a purely statistical phenomenon at work. Rising incomes lead to a rightward shift in the frequency distribution of the population across

[7] Growth in NNP per capita rather than GDP per capita is the relevant variable when considering the impact of growth on poverty.

income levels, and if a large proportion of the population is initially concentrated just below the poverty line, then the percentage below the poverty line will decline rapidly initially, followed by a slower response as the hump of the distribution moves above the poverty line.

Assessments of changes in poverty over time are affected by data problems and scholars have come to opposite conclusions. Some have argued that the official estimates present an unduly pessimistic picture because the estimates of personal consumption which emerged from the surveys are much lower than those implied by the national accounts.[8] Similar underestimation is found in most countries but in the Indian case, the discrepancy has also been widening sharply over time: the survey estimates were only about 20% lower in the 1960s but were 50% lower in 2004/5! If the consumption estimates from the surveys are adjusted upward in line with the national accounts, that will obviously reduce the extent of poverty and since the discrepancy has been increasing over time, the adjustment would also show poverty declining more sharply over time.[9] Bhalla (2004) estimates poverty by making a limited adjustment to the National Sample Survey (NSS) data after 1987 to limit the proportional discrepancy to the level reached in 1987. This limited adjustment results in poverty declining from 42.8% in 1983 to 9.3% in 2004/5.

Others have argued that the official estimates actually understate poverty because the poverty line is based on the per capita consumption level which in 1973/4 was associated with a food basket which provided a sufficient intake of calories per person. Although this 1973/4 poverty line is adjusted for inflation over time, it is argued that changes in the consumption basket have broken the link with adequacy of calorie intake. Expenditure on transportation has increased which may reflect the rising cost of commuting to work, and increased consumption on this account cannot be viewed as welfare-enhancing. Relative food prices have also increased, which may have reduced food intake at the same level of real per capita consumption. If the poverty line is adjusted upward to reflect these considerations the level of poverty would be higher. A separate case can also be made for revising the poverty line upward to reflect the fact that per capita incomes have increased. The need for such an adjustment was clearly envisaged in Planning Commission (1962).[10]

[8] There are conceptual reasons why survey estimates of consumption will be lower than the national accounts estimates because they do not include various types of institutional consumption, e.g. consumption in prisons or charitable institutions.

[9] Earlier, official estimates of poverty were derived by the Planning Commission after adjusting the NSS data proportionately upward to align them with the national accounts, but this practice was changed to using unadjusted data.

[10] The issue of whether the poverty line needs to be revised has been referred to an Expert Group set up by the Planning Commission and the report of the group is expected shortly. Such considerations have added contemporary relevance given movements in food prices in international markets. Domestic food prices have not increased correspondingly in India thus far, but a rise in relative food prices is likely over the next two years.

Table 17.4. Health and nutrition indicators, India, 1992/1993 to 2005/2006

	1992/3 (NHFS-1)	1998/9 (NHFS-2)	2005/6 (NHFS-3)
Children's health and nutrition[a]			
Percentage of children:			
Stunted	–	51.0	44.9
Wasted	–	19.7	22.9
Underweight	(51.5)[b]	42.7	40.4
Fully vaccinated	35.4	42.0	43.5
Safe delivery			
Percentage of women who:			
Benefit from institutional delivery	26.1	33.6	40.8
Benefit from safe delivery[c]	35.1	42.4	48.8

[a] Estimates of child malnutrition for 1998/9 and 2005/6 have been revised to reflect the new standards established by WHO using an international reference population. Estimates for 1998/9 and 2005/6 using the old standard were 47% and 45.9% respectively.

[b] The figure of 51.5% for 1992/3 for underweight children relates to the old standard and is therefore not comparable.

[c] Safe delivery = births in institutions or with presence of skilled birth attendant.

Source: *National Family Health Survey (NFHS-3) 2005–06, Key Findings* (Deonar, Mumbai: International Institute for Population Sciences, 2006).

III.2 Other Indicators of Deprivation

Poverty can also be measured in terms of deprivation in relation to health, nutrition, education, etc., and this provides a different picture from poverty measured in terms of the consumption poverty line. Available data, some of which are presented in Table 17.4, confirm that the percentage of the population deprived of essential services is significantly larger than the estimates of poverty.

The percentage of children suffering from severe malnutrition, for example, was 40.4 in 2005/6, significantly higher than the percentage of the population below the poverty line. The level of child malnutrition has declined over time but only relatively slowly especially after 1998/9 (see note *a* of Table 17.4), even though faster GDP growth might have been expected to accelerate the rate of decline.

The data on vaccination of children highlights a similar lack of access to health-related services. The percentage of children fully vaccinated was 35.4 in 1992/3 and increased to 42 in 1998/9, but thereafter increased only marginally to 43.5 in 2004/5. The fact that more than half of the children were left unvaccinated in 2004/5 certainly points to a major failure of public health delivery in an area which is critical and also relatively low in cost.

The percentage of women benefiting from safe deliveries tells a similar story: it increased from 35.1% in 1992/3 to 48.8% in 2005/6, suggesting significant improvement, but nevertheless more than half the births in India take place without the benefit of a skilled birth attendant. The comparable figure for China is 97%; Indonesia 72%; Thailand 99%; and Vietnam 85%.[11] The percentage of ever-married Indian women in the 15–49 age group who are anemic actually increased from 51.8% in 1998/9 to 56.2% in 2005.

Education-related indicators present a better picture. Illiteracy, as recorded in the Census, has been falling steadily. Even so, it was 39% in 2001 and is probably close to 30% today, whereas in South-East Asia, it had dropped to 20% by 1980. Enrollment in primary schools presents a fairly positive picture, with a net enrollment rate of 96% in 2005, reflecting the success of government programs aimed at expanding primary school infrastructure. However, high enrollment rates notwithstanding, 52% of children drop out before completing primary school, and as many as 63% do not complete secondary schooling, indicating a high degree of effective excluson. For children attending school, the quality of teaching in the public sector schools in many parts of the country is seriously deficient. Widespread teacher absenteeism, inadequate teacher training and outmoded pedagogy are major problems. Low-quality education in the public school system is leading many parents to shift children to private schools by paying fees which many can ill afford.

The children excluded from adequate education are predominantly from lower-income groups and belong to the scheduled castes and scheduled tribes, and the problem is more acute in rural areas. Unless access to proper education is radically improved, the system will perpetuate inequality of opportunity, denying the poor and excluded groups, especially in rural areas, the chance to benefit from the many opportunities created by rapid growth. Improving access to quality education for the poor is obviously crucial for reducing inequality in opportunity, which is in some ways more important than reducing inequalities of outcomes.

III.3 Inequality and Regional Imbalance

Concern about inequality surfaces in different forms: inequality in the distribution of incomes, rural–urban differences, inequality across states and inequality within states. Changes in inequality in India have been traditionally discussed on the basis of data on the distribution of consumption. The available data show no clear trend (Table 17.5.). The Gini coefficients of both urban and rural consumption increased in the 1970s up to 1983 and then declined up to 1993/4, after which

[11] This is undoubtedly one of the major reasons why the maternal mortality rate in India was 301 in 2001/2 per 100,000 births, compared with 56 in China, 230 in Indonesia, 130 in Vietnam and 44 in Thailand.

Table 17.5. Measures of inequality, India, 1973/1974 to 2004/2005

	1973/4	1977/8	1983	1993/4	2004/5
Gini coefficient of distribution of consumption					
Rural	0.28	0.34	0.30	0.28	0.30
Urban	0.30	0.34	0.33	0.34	0.37
Rural–urban ratio of mean consumption					
Current prices	0.75	0.72	0.69	0.61	0.53
Constant prices	0.63	0.65	0.65	0.61	0.58

Source: Estimated from National Sample Survey on consumer expenditure, various rounds.

they increased marginally. Even so, the Gini coefficient in 2004/5 was at about the same level for urban areas as in 1977/8 and for rural areas it was actually lower.

An aspect of inequality much discussed in recent years is the balance between rural and urban areas as measured by the ratio of per capita consumption in rural areas to per capita consumption in urban areas. Table 17.5 shows that this ratio has indeed swung against rural areas in the post-reform period, but this is only a continuation of a trend that was evident even before the reforms. The ratio declined steadily from 0.75 in 1973/4 to 0.69 in 1983, and then to 0.61 in 1993/4 and to 0.53 in 2004/5. The trend in constant prices is somewhat different, the ratio moving marginally in favor of rural areas up to 1983, after which it moved against rural areas. It is interesting to note that the fall in the rural–urban ratio in the post-reform period in constant prices is less than that in current prices.

Inter-regional inequality, or more specifically, inequality across States, is a dimension that has become particularly important because of the differing impact of reforms on the growth rates of different states. This aspect acquires special significance in a political environment in which government at the national level consists of coalitions that include state-based parties. A growth strategy which generates widely different performance across states will obviously generate greater political criticism of the reforms in those states that appear to lose out.

Table 17.6 presents the rates of growth of Gross State Domestic Product (GSDP) in each of the 14 major states for the 1980s and for the period after the economic reforms (1991/2 to 2004/5). The extreme view, often articulated in the popular press, that the "rich" states are getting richer and the poor states are getting "poorer" is simply not true. The richest states have certainly not grown fastest. Punjab, for example, which was until recently the richest state in terms of per capita GSDP, decelerated in the post-reform period, growing significantly more slowly than the national average. Some of the middle-ranking states, such as Gujarat, Maharashtra, Karnataka, Kerala and West Bengal, have accelerated compared to their pre-reform

Table 17.6. Growth rates of major states, India, 1981/1982 to 2004/2005

State	Growth in GSDP (% per annum)	
	1981/2 to 1990/1	1991/2 to 2004/5
Andhra Pradesh	6.4	5.8
Assam	4.2	3.5
Bihar	4.9	3.4
Gujarat	5.9	7.3
Haryana	6.6	5.5
Karnataka	5.2	7.0
Kerala	3.7	6.2
Madhya Pradesh	5.2	4.3
Maharashtra	6.1	6.4
Orissa	3.3	5.4
Punjab	5.3	4.5
Rajasthan	8.1	5.6
Tamil Nadu	5.7	5.6
Uttar Pradesh	5.1	3.8
West Bengal	4.4	6.9

Source: Estimated from data on GSDP obtained from the Central Statistical Organisation; available at <http://mospi.nic.in/13_gsdpcnst8081.ser.htm> and <http://mospi.nic.in/7_gsdp_const_9394ser.htm>.

position. No state has experienced actual impoverishment in the shape of a decline in per capita GSDP. It is true, however, that the acceleration in growth observed for the economy as a whole has not been experienced by all states equally. In fact, some of the poorest states, e.g. Uttar Pradesh, Rajasthan, Bihar and Madhya Pradesh, have decelerated in the post-reform period. Since these states account for about 30% of the total population of the country, and they have 149 Members of Parliament out of a total of 552 in the Lok Sabha (the Lower House of Parliament), the experience of these states is bound to color political perceptions. It should be noted, however, that some of the lower-income states have done relatively well in terms of growth. Orissa, for example, has accelerated its growth, though it is important to note this has not led to a significant reduction in poverty (see Table 17.2), probably because growth in Orissa has been led by mineral extraction which is often distributionally inequitable.

The impact of differences in growth rates across states on inter-state inequality can be measured by constructing a Gini coefficient for the country as a whole on the assumption that all persons in each state have the per capita GSDP of the state. This measure reflects pure inter-state inequality, ignoring inequality within each state, and it shows that inter-state inequality increased modestly through the 1980s, with

the Gini coefficient rising from 0.15 in 1980/1 to 0.17 in 1990/1, followed by a sharp increase thereafter to 0.23 in 2004/5.

A new dimension of concern regarding regional balance is that states such as Gujarat, Maharashtra and Karnataka, which have benefited from the economic reforms, have areas within the state which do not appear to have shared in the process. The growth of Maharashtra, for example, is largely a reflection of growth in the coastal regions, with interior Maharashtra benefiting far less. These backward areas within states have begun to demand a better deal, and this has led the central government to initiate a Backward Regions Grant Fund which provides assistance to 200 identified backward districts in support of district-specific plans which are to be evolved at the district level, with the involvement of local elected bodies. Assistance provided from this window is in addition to the normal central assistance given to state governments, which is itself based on a formula which gives greater weight to poorer states. It is too early to judge whether this initiative has been effective.

III.4 Employment Generation

A common criticism of post-reform performance in India is that the growth generated has been of the "jobless" variety. Total employment growth did indeed slow down from 2.6% per year in the ten years from 1983 to 1993/4 to 1.25% per year in the immediate post-reform period 1993/4 to 1999/2000, after which it accelerated again to 2.6% per year from 1999/2000 to 2004/5. However, since the labor force grew faster than employment in both periods, the unemployment rate increased from 6.1% in 1993 to 7.3% in 1999/2000 and further to 8.28% in 2004/5. The increase in the unemployment rate is clearly a worrying factor, though it is worth noting that it remains below the level of 9.2% in 1983.

The acceleration in employment growth in the most recent period is potentially a positive development. Rangarajan *et al.* (2007) have argued that if the estimated elasticity of total employment with respect to GDP of 0.48 is applied to the 9% GDP growth now targeted, employment growth will exceed 4% per year, which would eliminate unemployment by 2009. It can be argued that other factors will adjust so that unemployment will not actually decline to zero. Faster employment growth may lead to higher rates of labor force participation and therefore faster growth in labor supply, though there are limits to this supply response. More importantly, the tightening of the labor market will raise real wages, which could trigger technology shifts that should moderate employment growth. However, even if unemployment is not eliminated, it is very likely to decline, and the quality of employment can also be expected to improve in terms of both real wages and the terms of employment.

From a distributional perspective, the quality of employment is actually much more important than the volume of employment. This is evident from the fact that even though measured unemployment was 8.3% in 2004/5, poverty in that year was 27.5%. Employment alone is clearly insufficient to escape from poverty; what is needed therefore is more high-quality employment, which in turn requires faster employment growth in the organized sector. Unfortunately, the data available on organized sector employment are not very reliable. Data from the Directorate General of Employment and Training (DGE&T) shows that employment in the organized private sector increased at an annual rate of 0.44% per year from 1983 to 1994, accelerating to 0.58% per year from 1994 to 2005. However, employment in the organized public sector, which had grown at 1.53% per year from 1983 to 1994, actually declined at 0.7% per year from 1994 to 2005 and this decline more than offset the positive growth in the private sector. Total employment in the organized sector therefore declined at 0.31% per year over the period 1994–2005. These data are widely quoted but their reliability is questionable because reporting of employment in a significant segment of the private sector is voluntary and this data base therefore suffers widespread under-reporting.

The National Commission on Employment in the Unorganized Sector has provided an alternative estimate using the two most recent NSS surveys of employment and unemployment. Their estimates, reported in Planning Commission (2008), show that employment in the organized sector has increased at 2.9% per year between 1999/2000 and 2004/5, but this growth is mainly in the category of "informal workers", i.e. employees who work regularly in the organized sector but do not benefit from full social security provisions. The number of formal workers in the organized sector benefiting from full social security is stagnant, slightly better than the negative growth reported in DGE&T data but distressing nevertheless. This picture of stagnation in formal employment in the organized sector, with an expanding underclass of "informal" employees, should be a matter of concern. It points to the possibility that labor market rigidities may be discouraging expansion of high-quality employment.

IV. Some Critical Issues

India's performance in the post-reform period suggests that rapid growth, which eluded planners for so long, now appears to be a real possibility; but the associated distributional outcomes fall short of expectations. The challenge facing policy-makers is how to strengthen the momentum of growth while ensuring at the same time that growth is inclusive. The recently approved Eleventh Plan addresses this

challenge by calling for faster and more inclusive growth. It targets a growth rate of GDP of 9% per year from 2007/8 to 2011/12 and targets poverty reduction at 2% per year, more than twice the rate achieved in the recent past. It also specifies 26 other monitorable targets relating to improvements in employment, education, health, child nutrition and rural infrastructure. Some critical aspects of this strategy are discussed briefly in the rest of this section.

A stronger agricultural growth performance is critical for inclusiveness and especially for those living in rural areas. The Eleventh Plan has therefore targeted agriculture GDP growth of 4% per year. This is to be achieved through a revival of public investment in irrigation, better watershed management in rainfed areas, expanded investment in rural roads and rural electrification, combined with efforts to encourage diversification into horticulture, livestock, dairying and fisheries. Since these products are perishable, expansion also requires modernization of agricultural marketing and development of supply chains and logistics, all of which calls for a greater role for the private sector. However, achieving 4% growth in agricultural GDP will not by itself raise rural incomes per head sufficiently unless labor is also shifted out of agriculture and into productive work in the non-agricultural sector. This calls for faster growth of industry, especially of the more labor-intensive variety, to create sufficient employment in the non-agricultural sector.

Better physical infrastructure is a critical requirement even for maintaining the growth rate already achieved, and even more so for accelerating it. Investment in infrastructure is also crucial for achieving better regional balance, since many of the poorer parts of the country are excluded from the growth process precisely because infrastructure development in these areas has lagged behind. Better infrastructure will bring in its train an expanded flow of private investment, without which it is difficult to imagine an improvement in regional balance. The Plan proposes to increase investment in infrastructure (electric power, roads, railways, ports, airports and telecommunications) from about 5% of GDP in 2006/7 to 9% by 2011/12. This is to be achieved by a combination of increased public investment and increased private investment in infrastructure through various types of public–private partnership. Several policy initiatives have been taken to attract private investment into each of the infrastructure sectors listed above and the good news is that private investments have begun to flow into these sectors. However, it remains to be seen whether private investment can be scaled up to the level required. Much will depend upon the credibility of the policy and regulatory environment for each sector and also, of course, on the state of the global economy.

Health and education are two critical areas where the deficiencies which exist at present can only be removed if the government plays a much larger role. The Eleventh Plan proposes to raise total public expenditure on health from 1% of GDP at the start of the Plan period to 2% at the end, and in the case of education from around 3.8% to 5% of GDP. In both areas, the increase in public resources is only a first step. It must be accompanied by reforms in the system of public service

delivery to improve outcomes (we return to this issue later in this section). In addition to general education it is also necessary to focus on skills development. The acceleration of growth in recent years has already produced skills shortages, and a substantial effort is needed to quickly produce the skills needed by a rapidly growing economy. In all these areas, there is scope for public–private partnership, which could help to leverage limited public resources, and this must be exploited effectively.

Infrastructure development combined with an expanded supply of skilled labor will help to accelerate industrial growth, which should help to create much-needed employment. This needs to be supported by initiatives to introduce greater flexibility into Indian labor laws. The most important constraints in this context are the provisions of the Industrial Disputes Act 1947, which require that decisions to retrench labor or to close down an industrial unit require prior government permission, which is rarely given. While other elements of labor legislation provide legitimate protection to labor, these provisions limit flexibility unduly and discourage investment and expansion of employment in labor-intensive industries. The net effect is to discourage expansion in precisely those industries which require relatively limited skills and which could most easily help to shift surplus labor out of agriculture. China has been remarkably successful in achieving this transformation by expanding its share in world markets for simple manufactured products. India could compete effectively with China for these markets if it had more flexibility than at present.

The growth strategy envisaged here requires a substantial increase in government investment in both economic and social infrastructure. These resources have to come from a combination of greater efficiency in mobilizing tax revenues, an increase in user charges where possible, and a rationalization of untargeted subsidies whose proliferation has been crowding out productive public investment. Fortunately, the experience on tax revenue mobilization has been very good, with a steep increase in the ratio of total taxes to GDP in recent years. Rapid growth of GDP combined with continued reform of the tax system will make a major contribution to generating the resources needed for programs aimed at making growth more inclusive. The proposed merger of the central and state indirect taxes into a unified goods and services tax, which is expected from April 2010, is a particularly important reform in this context. However, success in raising user charges and tackling subsidies remains elusive. The provision of free or virtually free electric power to farmers, combined with over-subsidization of fertilizers, not only absorbs a large volume of resources which could be more productively devoted to investment in irrigation and watershed management; it also distorts prices, leading to overuse of ground water and excessive use of fertilizers, which contributes to soil damage and lowering of the water table. Yet another subsidy that has grown in scale is the subsidy on petroleum products, especially kerosene and LP gas, because of the

failure to pass on the recent sharp increase in world prices of petroleum products. Unless the scale of these subsidies is reduced, it will not be possible to meet many of the resource requirements of inclusive growth.

The search for inclusive growth has highlighted the need to create a minimal system of social security for the poorer sections of the community. Several important steps have been taken in this area. A new National Rural Employment Guarantee Act has been passed under which every rural household has a statutory right to demand up to 100 days of manual labor at the notified minimum wage for casual labor. This not only supplements the income of agricultural laborers; in areas of significant underemployment, it also helps to tighten rural labor markets, raising wages generally. The employment provided is on asset-creating public works chosen by the *gram sabha* (village assembly). Watershed management and water conservation works are given priority, as these are most likely to increase land productivity, with subsequent income and employment benefits. The early experience is encouraging. Though there are examples of misuse and leakage, these seem to be much fewer than in earlier employment programs and beneficiaries in many areas are reporting high levels of satisfaction. Another major social security initiative is a medical insurance scheme covering hospital expenses for those below the poverty line, in which the entire premium is paid by the central and state governments (75% and 25% respectively). When fully deployed this program will extend medical insurance to nearly 300 million people, protecting them from hospital expenses that are often crippling. The insurance enables the insured to seek treatment in participating hospitals, which include both the public and the private sector, thus introducing an element of competition. Finally, an old-age pension scheme, which previously covered only "destitutes", has been expanded to cover all those above age 65 in households below the poverty line with a modest old-age pension. Taken together these programs constitute a significant improvement in social security for persons in the unorganized sector.

A major institutional challenge is the need to design systems which can effectively deliver essential services such as health, education, clean drinking water, sanitation, etc. to the mass of the people. Experience shows that existing systems of public service delivery, staffed by employees of the state government reporting vertically to departments in the state capital, are not effective. It is very difficult to ensure even minimum levels of accountability in terms of regular attendance by teachers in schools and health workers in clinics, especially in rural areas. It is possible that better results would be achieved if local governments were empowered to enforce accountability of staff engaged in service delivery at the local level. A major step in this direction was the amendment of the Constitution in 1993 to establish directly elected local bodies in each village and municipality, with a tiered system in which directly elected members of a village council (*panchayat*) in turn elect representatives for the block- and district-level government. This structure is now

in place, with about 3 million people elected to village councils across the country, one third of whom have to be women. However, these bodies also have to be effectively empowered in terms of devolution of functions, resources and control over functionaries, and progress in these areas remains limited. There are states where it is working well but there are many areas where much more remains to be done, including especially capacity building at the local level. Unless these tasks are carried out, there is a serious danger that additional public sector resources devoted to programs focused on distributional objectives will not be effectively employed.

These are some of the most immediate challenges to making growth more inclusive. There are others, relating to management of energy and water requirements, urbanization and climate change, which also need to be addressed as part of any long-term growth strategy. The approach to these problems is spelt out in Planning Commission (2008). If the system is able to address these issues as effectively as it has addressed the constraints on growth, India has a good chance of achieving the objective of faster and more inclusive growth.

References

Ahluwalia, Isher (1985), *Industrial Growth in India: Stagnation since the Mid-Sixties* (New Delhi: Oxford University Press).

Ahluwalia, Montek S. (2002), "Economic Reforms in India since 1991: Has Gradualism Worked?", *Journal of Economic Perspectives*, 16: 67–88.

Bhagwati, Jagdish N., and Desai, Padma (1970), *India: Planning for Industrialisation. Industrialisation and Trade Policies since 1951* (Delhi: Oxford University Press).

Bhalla, Surjit S. (2004), "Poor Results and Poorer Policy: A Comparative Analyses of Estimates of Global Inequality and Poverty", *CESifo Economic Studies*, 50: 85–132.

Guha, Ramachandra (2007), *India after Gandhi: The History of the World's Largest Democracy* (London: Macmillan).

Ozler, Berk, Datt, Gaurav, and Ravaillon, Martin (1996), *A Database on Poverty and Growth in India* (Washington, DC: World Bank).

Planning Commission (1962), "Perspective of Development: 1961–1976. Implications of Planning for a Minimum Level of Living", repr. in Srinivasan and Bardhan (1974).

—— (1964), *Report of the Committee on Distribution of Income and Levels of Living: Part I* (New Delhi: Planning Commission, Government of India).

—— (2008), *Eleventh Five-Year Plan. Volume I: Inclusive Growth* (New Delhi: Planning Commission, Government of India).

Rangarajan, C., Kaul, Padma Iyer, and Seema (2007), "Revisiting Employment and Growth", *ICRA Bulletin*, September: 57–68.

Rodrik, Dani, and Subramanian, Arvind (2004a), *From "Hindu Growth" to Productivity Surge: The Mystery of the Indian Growth Transition*, NBER Working Paper 10376 (Cambridge, Mass.: National Bureau for Economic Research).

—— —— (2004b), *Why India Can Grow at 7 Percent a Year or More: Projections and Reflections*, IMF Working Paper 04/118 (Washington, DC: International Monetary Fund).

Shah, K. T. (ed.) (1949), *Report: National Planning Committee*, 1st edn (Bombay: Vora & Co. Publishers Ltd).

Srinivasan, T. N., and Bardhan, Pranab (1974), *Poverty and Income Distribution in India* (Calcutta: Statistical Publishing Society).

Tendulkar, Suresh D. (1974), "Planning Models for Growth and Distribution in India: An Assessment", in Srinivasan and Bardhan (1974).

CHAPTER 18

ECONOMIC REFORMS, POVERTY AND INEQUALITY IN CHINA AND INDIA

PRANAB BARDHAN

I

In various writings over the years Amartya Sen has shown an acute interest in the comparative performance of China and India, the two largest countries of the world, both with poor, agrarian economies and rich, ancient civilizations. While high economic growth in both countries in recent decades has attracted a lot of attention, particularly in the international financial press, Sen's interest has usually been more in issues relating to poverty, inequality and gender inequities in general, and health and education matters in particular, in both countries. At various times his thoughts on the better achievements of China in literacy and life expectancy than India, on the era of market reform in China coinciding with the decline of basic public health services there, on democratic Kerala's demographic policies (significantly linked with female education) achieving more than China's coercive one-child policy, on the "missing women" in both China and India attributable

to gender discrimination, on China faring much better than India in relieving "endemic hunger" while democracy and the freer media in India help avert catastrophes like the Great Famine in China in the early 1960s—these have all become hallmarks of development folklore. Even in his piece on the cultural and intellectual links between China and India in the first millennium (Sen 2004), he does not forget to emphasize the interest of Chinese visitors to ancient India in the public health care system and the tradition of democratic dialogue and public discussion that Buddhism carried to China. In this chapter I shall follow in this comparative tradition and focus on the impact of economic reform in the two countries on poverty and inequality and their interaction in the political process.

II

In the last quarter century a great many reforms have taken place in both China and India—in market liberalization and macroeconomic policy, trade and industrial policy, tax and financial policy, privatization and deregulation. In China the reform in agriculture, with de-collectivization and the shift to the household responsibility system around 1979, along with a rise in agricultural procurement prices, led to a large growth in the agricultural sector. According to the estimate of Lin (1992), agricultural output grew at 7.1% per year on an average during 1979 to 1984 compared to 2.7% during 1970 to 1978. As much of the extreme poverty was in the rural sector, this led to a large fall in poverty. If one takes the admittedly crude World Bank poverty line of $1 a day per capita (at 1993 purchasing power parity), the proportion of people below that poverty line in China fell from 64% in 1981 to 29% in 1987 (see Table 18.1).[1]

This is a dramatic decline in a short span of years. Part of this decline may not be real: poverty may have been somewhat overestimated in 1981, as the official price deflator used for this estimation may not have been adequate for rural areas before 1985. But a large part of this decline may be genuine, and not due just to the large

[1] One dollar per capita per day is an arbitrary cut-off line with very little link to any estimate of the basic needs of the poor. Also, the calculations of purchasing power parity are not sensitive to the different subsistence needs of the poor in different countries and cultural contexts. I nonethless use this here because we do not have access to any alternative time series of poverty data which we can compare across countries. We should, however, note that the qualitative results about change over time stated in the text will be the same even with the national estimates using local poverty lines in China and India.

The World Bank is soon to release revised estimates of Chinese and Indian poverty on the basis of their new purchasing power parity numbers. The estimates reported in this chapter are based on their earlier PPP numbers.

Table 18.1. Population in poverty (less than $1 a day per capita in 1993 PPP), China and India, 1981–2004

	1981	1984	1987	1990	1993	1996	1999	2002	2004
Percentage of population									
China	63.76	41.02	28.64	32.98	28.36	17.37	17.77	13.79	9.90
India	54.31	49.50	45.88	44.29	42.13	40.61	38.76	37.49	35.78
Number of people (millions)									
China	633.66	425.27	310.43	374.33	334.21	211.44	222.78	176.61	128.36
India	381.67	371.08	366.45	376.21	378.93	385.30	387.27	393.16	386.37

Source: Chen and Ravallion (2007).

rise in the agricultural growth rate during this short period, but also, perhaps, to the fact that de-collectivization was associated with what may have been one of history's most egalitarian land redistribution exercises, with every rural family getting an equal piece of land (subject to differences in family size and regional average), thus giving a floor to the household income of the poorest people.

In the non-agricultural sector in China, the new contract system with firms and decentralization, which allowed the local governments and collectives to be "residual claimants" of the enterprises they ran (township and village enterprises) and to compete with one another, led to a phenomenal advance, particularly in rural industrialization. Foreign business was allowed to operate freely in the Special Economic Zones. Governance in the state-owned enterprises was restructured and corporatized and made more profit-oriented; in the initial period they faced more competition from collective and other enterprises, and more recently, many have been privatized. As Table 18.2 shows, total factor productivity in industry grew at an annual average of 3.1% in 1978–93 and at double that rate in 1993–2004.

In India the reforms of the 1990s seem to have been associated with a more vigorous and competitive corporate sector, but most of the economy is outside the corporate sector. In Table 18.2 one notes a rise in the annual growth of total factor productivity in industry, from 0.3% in 1978–93 to 1.1% in 1993–2004. The more significant rise in India is in the service sector; growth in total factor productivity in that sector rose from an annual average of 1.4% in 1978–93 to 3.9% in 1993–2004. The Indian growth process has been described as a service-sector-led growth, whereas in China it has been more manufacturing-centered (note, however, in Table 18.2 that in the first period, 1978–93, even service sector total-factor productivity grew faster in China than India). One immediately thinks of the widely acclaimed performance of Indian software and other IT-enabled services. But it seems that not all of the growth in the economy's service sector during the period 1993–2004 can be explained by finance, business services or telecommunications, where reform may have made a difference.

Table 18.2. Sources of growth by major sector, China and India, 1978–2004 (annual percentage rate of change)

Period	Output	Employment	Output per worker	Contribution of			
				Physical capital	Land	Education	Factor productivity
Agriculture							
1978–2004							
China	4.6	0.3	4.3	2.3	0.0	0.2	1.8
India	2.5	1.1	1.4	0.4	−0.1	0.3	0.8
1978–93							
China	5.2	0.9	4.3	2.5	−0.2	0.2	1.8
India	2.7	1.4	1.3	0.2	−0.1	0.2	1.0
1993–2004							
China	3.7	−0.6	4.3	2.1	0.2	0.1	1.8
India	2.2	0.7	1.5	0.7	−0.1	0.3	0.5
Industry							
1978–2004							
China	10.0	3.1	7.0	2.2		0.2	4.4
India	5.9	3.4	2.5	1.5		0.3	0.6
1978–93							
China	9.3	4.4	4.9	1.5		0.2	3.1
India	5.4	3.3	2.1	1.4		0.4	0.3
1993–2004							
China	11.0	1.2	9.8	3.2		0.2	6.2
India	6.7	3.6	3.1	1.7		0.3	1.1
Services							
1978–2004							
China	10.7	5.8	4.9	2.7		0.2	1.9
India	7.2	3.8	3.5	0.6		0.4	2.4
1978–93							
China	11.3	6.5	4.7	1.8		0.2	2.7
India	5.9	3.8	2.1	0.3		0.4	1.4
1993–2004							
China	9.8	4.7	5.1	3.9		0.2	0.9
India	9.1	3.7	5.4	1.1		0.4	3.9

Source: Bosworth and Collins (2007).

Table 18.3 shows that a large part of the growth in the Indian service sector, at a rate higher than in manufacturing, was in the traditional or "unorganized sector" services, which even in the last decade formed nearly two-thirds of service sector output. These are provided by tiny enterprises, often below the policy radar, unlikely to have been directly affected substantially by the regulatory or foreign trade policy reforms. It is a matter of some dispute how much of the growth in traditional services (mostly non-traded) is explained by the rise in service demand

Table 18.3. Growth in components of service sector, India, 1960–2004 (%)

Period	Modern services					Traditional services				Services less dwellings
	Total	Communications	Finance	Business services	Education and medical	Total	Trade	Transportation	Other services	
Share of total output in services										
1960/1	19	2	6	1	10	81	40	14	27	100
1980/1	22	3	7	1	11	78	37	16	24	100
1993/4	31	3	14	2	12	69	34	14	21	100
1999/2000	35	6	14	4	12	65	33	12	19	100
2004/5	40	11	12	5	11	60	33	11	16	100
Annual rate of change										
1960–80	5.7	6.9	5.9	3.4	5.5	4.6	4.5	5.6	4.3	4.9
1980–93	9.0	7.1	12.3	9.8	6.6	5.4	5.6	5.4	4.9	6.3
1993–9	12.6	20.3	9.3	28.0	10.6	8.9	9.8	7.5	8.6	10.1
1999–2004	10.5	23.8	5.7	11.4	7.1	6.5	7.9	5.7	4.3	8.0
Contribution to total services growth										
1960–80	1.1	0.1	0.3	0.0	0.5	3.8	1.8	0.8	1.2	4.9
1980–93	2.0	0.2	0.9	0.1	0.7	4.2	2.1	0.9	1.2	6.3
1993–9	3.9	0.7	1.3	0.5	1.2	6.2	3.3	1.1	1.8	10.1
1999–2004	3.7	1.3	0.8	0.5	0.8	4.2	2.6	0.7	0.8	8.0

Source: Bosworth, Collins and Virmani (2007).

in the rest of the economy (including increased outsourcing by the manufacturing firms which formerly used to supply those services in-house), and how much is a statistical artifact, as the way the output is measured in these traditional services has been rather shaky all along, both with respect to the estimates made for some benchmark years and for the methods of updating them, particularly in the absence of an appropriate price index for the service sector.[2]

So the link between economic reforms and growth in the whole economy is not yet clearly established, even though it is very likely that the reduction in controls and regulations and the increased leeway of market discipline and the forces of competition may have unleashed entrepreneurial energies in both the formal and informal sectors. (I would also like to speculate that the concurrent social changes in India, namely the political rise of hitherto subordinate social groups after many centuries of social oppression, may also have played some role in this unleashing of energies.)

III

It is often claimed, both in the media and academia, that it was global integration which brought down the extreme poverty that had afflicted China and India over many decades. While expansion of exports of labor-intensive manufactures lifted many people out of poverty in China in the last decade (albeit not in India, where exports are still mainly skill- and capital-intensive), the more important reason for the dramatic decline of poverty over the last three decades may actually lie elsewhere, as we have already indicated above. Table 18.1 suggests that two-thirds of the total decline in the numbers of poor people (below the poverty line of $1 a day per capita) in China between 1981 and 2004 had already happened by the mid-1980s, *before* the big strides in foreign trade and investment in China in the 1990s. Much of the extreme poverty was concentrated in rural areas, and, as we noted above, its large decline in the first half of the 1980s is perhaps mainly a result of the spurt in agricultural growth following de-collectivization, land reform, and upward readjustment of farm procurement prices—mostly internal factors that had very little to do with global integration.[3] Ravallion and Chen (2007: 3) conclude from

[2] For a detailed discussion of the problems in measuring output in the service sector, both in the corporate and the unorganized subsectors, see the symposium in *Economic and Political Weekly*, 42(17), 15–21 September 2007.

[3] In the 1980s there was some trade expansion, for example the export-to-GDP ratio went up from about 7% to 12% in 1989. But labor-intensive manufactures were still not very important in Chinese export; in the first half of the 80s, minerals and other natural-resource-intensive products formed a substantial fraction of exports. In 1985, for example, the largest single export was petroleum. The

their analysis that "the score-card for trade reform is blank: we find no evidence that greater external openness was poverty reducing". (To settle these issues, of course, one needs causal models which are yet to be tested in the literature.)

In India, National Sample Survey (NSS) data suggest that the rate of decline in poverty has, if anything, somewhat slowed down from 1993 to 2005, the period of intensive opening up of the economy, compared to the seventies and eighties. This may not be unconnected with the fact that agricultural output (and total factor productivity) grew at a slower rate in the last decade compared to the earlier decade (see Table 18.2). This may be largely on account of the decline in public investment in rural infrastructure (such as irrigation, roads and prevention of soil erosion), which has little to do with globalization. There has also been a decline in the rate of growth of real wages in the period 1993–2005 compared to the previous decade, 1983–93. Besides, we should recognize that the private consumer expenditure data of the NSS that are used in poverty estimates do not capture the decline in environmental resources (like forests, fisheries, grazing lands, and water, both for drinking and irrigation) on which the daily lives and livelihoods of the poor depend.

Global integration does not seem to have helped some of the other non-income indicators, like those of health. The National Family Health Survey (NFHS) data show that some of India's health indicators are worse than those of Bangladesh (maternal mortality, infant mortality, child immunization rates, etc.), and even those of sub-Saharan Africa (percentage of underweight children), in spite of much higher growth rates in India than in those other countries. The percentage of underweight children (below age 3) is 46 in India, and about 30 on average in sub-Saharan Africa (8 in China). Take the case of Gujarat, one of the richest, highest-growth, and highest-reform states in India: the percentage of underweight children, which was already high (higher than sub-Saharan Africa), went *up* between NFHS II (1998/9) and NFHS III (2005/6).

Some disaggregated studies across districts in India have also found trade liberalization slowing down the decline in rural poverty.[4] Such results may indicate the difficulty of displaced farmers and workers in adjusting to new activities and sectors on account of various constraints (for example, constraints on getting credit or

mean tariff rate declined only slightly in the 80s, from 31.9% in 1980–3 to 29.2% in 1988–90. In any case the proportion of the labor force in manufacturing in this period was small, so the large poverty decline in the first half of the 80s is unlikely to be attributable to manufacturing exports. It is also worth noting that the percentage in poverty, after the sharp drop between 1981 and 1987, went up (or remained the same) between the years 1987 and 1993. This indicates that by 1987 the agricultural spurt had worked itself out and the effect of labor-intensive manufactures was still weak. It was only after 1993 that the poverty percentage again started declining and labor-intensive exports may have played a significant role in it, although even in this period one should not minimize the effect of largely domestic factors such as easier migration from rural areas and higher agricultural procurement prices.

[4] See, for example, Topalova (forthcoming). In unpublished comment T. N. Srinivasan has raised some doubts about the methods in this study.

information or infrastructural facilities like power and roads; or the large number of school dropouts; or labor market rigidities), even when new opportunities are opened up by globalization. This is in line with textbooks in international economics where it is emphasized that product-market liberalization need not be an improvement when there are severe distortions in input markets.

The Indian pace of poverty reduction has been slower than China's, not just because growth has been faster in China, but also because the same 1% growth rate reduces (or is associated with reduction in) poverty in India by much less. The so-called growth elasticity of poverty reduction is much higher in China than in India; this may have something to do with the different inequalities in wealth in the two countries (particularly in land and education). Contrary to common perception, these inequalities are much higher in India than in China. The Gini coefficient of land distribution in rural India was 0.74 in 2003; the corresponding figure in China was 0.49 in 2002.[5] India's educational inequality is one of the worst in the world: according to the World Bank (2006: Table A4) the Gini coefficient of the distribution of adult years of schooling in the population, a crude measure of educational inequality, was 0.56 in India in 1998/2000, which is not just higher than China's 0.37 in 2000, but even higher than almost all Latin American countries (Brazil's is 0.39).

Across states in India, as Datt and Ravallion (2002) point out, the growth elasticity of poverty reduction depends on the initial distribution of land and human capital. Purfield (2006) indicates that in the period 1977–2001 this elasticity was quite low in high-growth states like Maharashtra and Karnataka, and high in states like Kerala and West Bengal. Similarly, when comparing states in China, Ravallion and Chen (2007) find that growth has more poverty-reducing impact in provinces that are initially less unequal.

IV

The link between economic reform and inequality in China and India is also not very clear. At least two major problems beset the analyst in this matter. One is that so many other changes have taken place in the last quarter century of reform, it is difficult to disentangle the effect of reform from that of other ongoing changes (such

[5] This, of course, does not correct for land quality. Land quality is partly taken into account in its valuation in India when land is included in the Assets and Liabilities Survey. According to this Survey by the NSS, the Gini coefficient of asset distribution in rural India was 0.63 in 2002, while the corresponding figure for China was 0.39 in the same year. For the Chinese estimate, see Li, Wei and Jing (2005). The Indian estimate is mine. The land Gini estimate for China cited in the text is from Khan (2004).

as technological progress—often skill-biased—demographic changes or macroeconomic policies). Secondly, in both countries there are reasons to suspect that economic inequality (or its rise) is underestimated because of the widely noted difficulty facing household surveys (in many countries) of large (and increasing) non-response by rich households. It is also difficult to compare China and India, as most of the inequality data cited in this context are usually for income inequality for China and consumption expenditure inequality for India (as the Indian NSS does not collect income data). These two disparate sources do show a rise in expenditure inequality in both countries in the last decade or so. But, as we have suggested, this rise may be an underestimate, and there is very little analysis as yet to show that this rise is primarily due to economic reform.

Even if economic reform were to be causally linked with higher growth, the link between growth and inequality is not always clear. In China, as Chaudhuri and Ravallion (2006) show, the periods of rapid growth have not necessarily brought more rapid increases in income inequality; the periods of falling inequality (1981–5 and 1995–8) had among the highest growth rates in average household income. In both countries, periods of high agricultural growth may have reduced overall inequality, and the recent decline in agricultural growth rates may have had some influence on the rising inequality.

For the urban sector in India, Figure 18.1 gives some data on change in the real wage rate for urban full-time employees in the last two decades by level of education, and shows a faster rise in the wage rate for those with higher education. According to estimates by the Asian Development Bank (2007), the Gini coefficient of average real wages of urban full-time employees in India went up from 0.38 in 1983 to 0.47 in 2004. This increase in wage inequality is consistent with the skill-intensiveness of Indian economic growth (which reforms may have played some role in) and the looming talent shortage of which the corporate sector is complaining.

In urban China also the rate of return to college (and greater) education compared to, say, high school education has more than doubled since the early 1990s. Again, in both China and India it is difficult to separate the effect of skill-biased technological progress from that of economic reform.

Since reform has been arguably more "urban-biased" in India and (at least since the 1990s) in China, we should look at the urban–rural disparity. The ratio of urban to rural mean income (or consumption) is higher in China than India, and it is rising in India, but the rise in disparity in China is attenuated once one takes into account the rural–urban differences in cost of living.[6] Figure 18.2 decomposes

[6] Another statistical problem is that the definitions of households in China as rural or urban are somewhat uncertain, because (1) of increasing reclassification of rural areas as urban in official data, and (2) the location of households by residence is different from that by registration (in view of the restrictions of the *hukou* system).

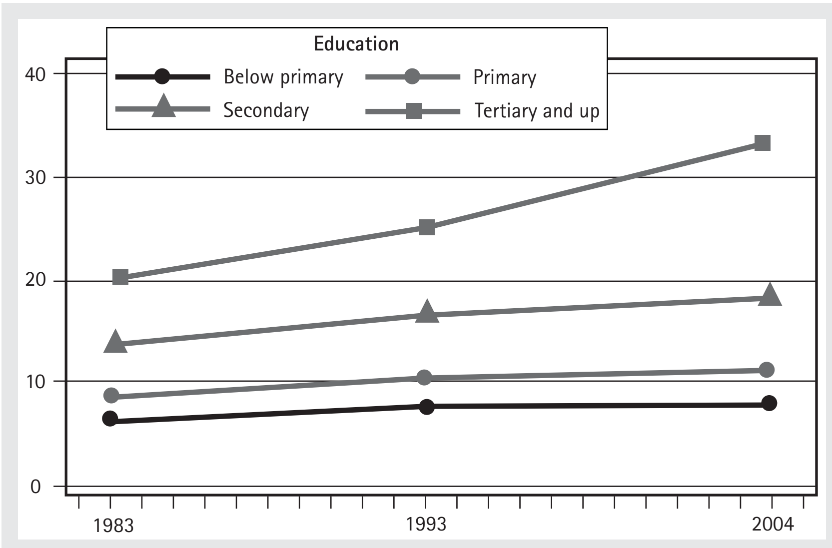


Fig. 18.1. Average weekly real wage of urban full-time employees by level of education, India, 1983–2004 (2002 US$)

Source: Asian Development Bank (2007).

national inequality in China into inequality within each of the rural and urban sectors and inequality between the rural and urban sectors. It seems that both intra-urban and urban–rural disparities contribute more to recent rises in income inequality.

In China the urban–rural disparity in social services has increased in the post-reform period, with a near-collapse of rural social services, whereas in urban areas social services, though weaker than before, still serve the majority of the resident (and working) urban population. An example from health care will serve: in 1985 the total number of technical medical personnel per 1,000 people was somewhat lower in the city than in the county; 20 years later, it was more than twice in the city.

With unfunded mandates from above (particularly in social services and pensions) and inadequate (and often regressive) central transfers and revenue-sharing, local public finances in China have been in bad shape, leading local officials in the recent past sometimes to impose arbitrary fees and "extra-budgetary" revenues, particularly in inland, agriculture-dominant areas.[7] In India, of course, delivery of

[7] The Chinese leadership at the top is now taking some remedial steps (e.g. rural tax reform in the last few years, a supplementary funding plan to offer free 9-year compulsory education in rural areas, and a partially subsidized medical insurance scheme to cover many rural areas). How effective and

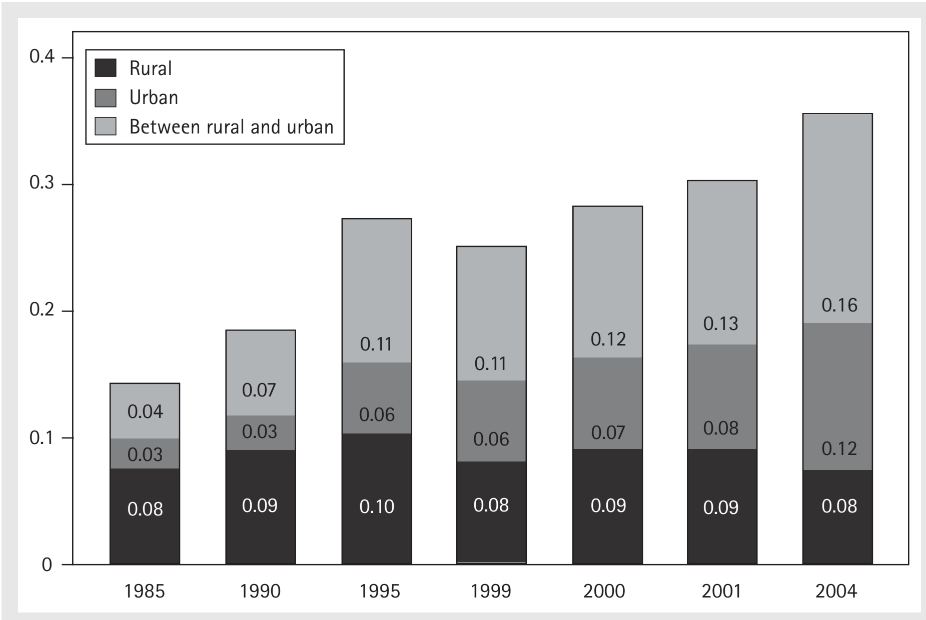


Fig. 18.2. Decomposition of Theil measure of national inequality in China, 1985–2004

Note: Inequality is measured in terms of the Theil index and ranges from 0 to 1.
Source: Asian Development Bank (2007).

public education and health has been dismal for many decades, as Amartya Sen has tirelessly pointed out for many years. It is not much of an exaggeration to say that China is now catching up with India in the matter of appalling public services in the rural sector. The urban population in China have also gained disproportionately from the privatization and marketization of housing since the mid-1990s.

The decline of rural health services in China (along with the one-child policy) may have had some effect on gender equity in life chances. The male–female sex ratio in children (below 6 years) is very high, at about 1.19 (1.08 in India). But one should add that female literacy and labor participation rates are substantially higher in China (the latter rate is above 70% in urban China, compared to 24% in urban India), so women in China have had much more opportunity to contribute to economic growth than in India.

Regional disparity in income (or consumption) is also more pronounced in China than in India. But over the last two decades China's backward regions have

adequate this will actually be when implemented remains to be seen. An unpublished 2005 survey at the Centre for Chinese Agricultural Policy indicates that while the tax burden on farmers fell, reduction of the local bureaucracy and the promised supplementary upper-level transfers have been inadequate, while the centralization of tax reform has reduced the incentives of local governments to serve social needs.

grown at rates almost comparable to its advanced regions, and regional earning disparities may be narrowing (though not yet disparities in per capita income). Initiatives like the Western Development Program supported large infrastructure building in the remote regions; and the mobility-restricting effects of the *hukou* system may be weakening. In any case Benjamin *et al.* (2005) have shown that the contribution of inter-provincial inequality to total inequality in China is smaller than is usually thought.

In India the poorer states (largely concentrated in the central and eastern regions) have grown much more slowly than the richer states (mostly in the west and the south), so relative inequality has increased. In general reform has advanced more smoothly in the west and the south of India, and better implementation of reform in a state may have gone hand in hand with better initial infrastructure. Of course, with the removal of industrial licensing, which ostensibly used to give some weight to regional backwardness, private capital will gravitate more to states where policies are business-friendly and infrastructure is better. Even in aspects of human development—for example, in child malnourishment—there is some evidence that the population-weighted coefficient of variation across states has increased between 1992/3 and 2005/6 (see Nair 2007).

In China the provinces with more global exposure and higher growth have not had larger rises in inequality. As Benjamin *et al.* (2005) note, while the Gini coefficient of income in coastal China went up from 0.35 in 1991 to 0.39 in 2000, the corresponding rise in the interior provinces was from 0.39 to 0.48. In the coastal provinces, more rapid job growth in the non-state sector helped reduce the urban–rural income difference there.

V

While reforms per se may thus have had a rather ambiguous effect on inequality, let us end with a general comment on (1) the link between initial inequality and the reform process and (2) the link between rising inequality and social discontent, which in turn affects the sustainability of the reforms. I believe that in the Indian situation of extreme social heterogeneity and income inequality in a contentious democratic framework, what happens to inequality is important for the success of the reform process, particularly because inequality makes the social and political environment quite conflict-ridden, and it is difficult in this environment to build consensus and organize collective action towards long-term reform and cooperative problem-solving. When groups do not trust one another to share the costs and benefits of long-run reform, there is an inevitable tendency to go for "bird-in-hand"

short-run subsidies and government handouts instead. This is in line with a large theoretical and empirical literature on the relation between inequality and collective action (see, for example, Baland 2006 and Bardhan 2005).

I believe this is not unconnected with the fact that reforms outside the corporate sector in India have been rather halting and hesitant, particularly reforms in the factor markets (land, labor, public inputs like electricity, etc.), which directly affect the lives and livelihoods of large numbers of people. In the CSDS National Election Survey data for 2004, two-thirds of the sample of about 23,000 respondents who had any opinion on the subject said that the reforms "benefit only the rich" or no-one at all. A strong majority of respondents say they are opposed to privatization and a reduction in the size of government (see Suri 2004). A similar survey by CSDS with a smaller sample in 2007 had similar results.

These attitudes, of course, partly reflect a failure of Indian reformers and politicians to explain the benefits of reforms to common people. But what financial columnists call anti-reform populism is also partly a product of the manifold inequalities and conflicts of Indian society. For example, the severe inequality in education (which, as we mentioned before, is much worse in India than in China and most Latin American countries), makes it harder for many workers to absorb shocks in the industrial labor market, since education and training could provide some flexibility in adapting to market changes.

In China the disruptions and hardships of restructuring under a more intense process of global integration were rendered somewhat tolerable in the eighties and nineties by the fact that China has had some kind of minimum rural safety net, largely made possible by an egalitarian distribution of land cultivation rights following the de-collectivization of 1979. In most parts of India there is no similar rural safety net for the poor. So resistance to the competitive process that market reform entails is that much stiffer in India. This is in line with a worldwide phenomenon: resistance to globalization is generally stronger in countries where social safety nets (particularly unemployment benefits and portable health insurance) are weaker (compare Scandinavian countries and the US in this respect).

Even though initial inequality may have been much lower in China, sharply rising inequality can be a major source of social and political discontent over time. Some have already linked this with the large number of incidents of unrest in different parts of the country, reported even in Chinese official police records. One should not, however, exaggerate the extent of inequality-induced discontent in rural and remote areas of China. Data from a 2004 national representative survey in China, carried out by Martin Whyte, a Harvard sociologist, and his team (reported in Feng 2007) show that those presumed disadvantaged in the rural or remote areas are not particularly upset by rising inequality.

This may be because of the familiar "tunnel effect" in the inequality literature attributed to Albert Hirschman: when you see other people prospering you are hopeful that your chance will come soon (just as you are more hopeful in a traffic

jam in a tunnel when the blocked traffic in the next lane starts moving); this is particularly so with the relaxation of restrictions on mobility from villages and improvement in roads and transportation. In any case, even in rural areas the average per capita household income increased at an annual rate of about 5% from 1991 to 2004. Even across expenditure groups, the bottom quintile in China experienced a significant 3.4% annual growth rate in mean per capita expenditure between 1993 and 2004 (the corresponding figure for the Indian bottom quintile is only 0.85%). Chinese leaders have also succeeded in deflecting much of the wrath towards corrupt local officials and in localizing and containing rural unrest. In any case, rural people in China are often more upset about forcible land acquisitions and toxic pollution than about relative inequality.

Of course, both China and India have done much better in the last quarter century, both in economic growth and poverty reduction, than in the last two hundred years. But in this chapter we have tried mainly to dispel some prevailing myths on the links between economic reforms, particularly those connected with global integration, and economic growth and improvement in the lives of the poor in these two countries. We have emphasized the various ambiguities and complications in these links and have also briefly commented on the tortuous relation between inequality, conflicts and the reform process itself.

References

Asian Development Bank (2007), *Key Indicators 2007: Inequality in Asia* (Manila: Asian Development Bank).

Baland, J.-M., and Platteau, J.-P. (2007), “Collective Action on the Commons: The Role of Inequality”, in J-M., Baland, P. Bardhan and S. Bowles (eds), *Inequality, Cooperation, and Environmental Sustainability* (Princeton, NJ: Princeton University Press).

Bardhan, P. (2005), *Scarcity, Conflicts, and Cooperation* (Cambridge, Mass.: MIT Press).

Benjamin, D., Brandt, L., Giles, J., and Wang, S. (2005), “Income Inequality during China’s Economic Transition”, in L. Brandt and T. Rawski (eds), *China’s Great Economic Transformation* (Cambridge: Cambridge University Press).

Bosworth, B., and Collins, S. M. (2007), *Accounting for Growth: Comparing China and India*, NBER Working Paper 12943 (Cambridge, Mass.: National Bureau for Economic Research).

—— —— and Virmani, A. (2007), “Sources of Growth in the Indian Economy”, in *India Policy Forum 2007* (New Delhi: Sage Publishers).

Chaudhuri, S., and Ravallion, M. (2006), “Partially Awakened Giants: Uneven Growth in China and India”, in L. A. Winters and S. Yusuf (eds), *Dancing with Giants: China, India, and the Global Economy* (Washington, DC: World Bank).

Dutt, G., and Ravallion, M. (2002), “Has India’s Post-Reform Economic Growth Left the Poor Behind?”, *Journal of Economic Perspectives*, 16: 89–108.

Feng W. (2007), "Boundaries of Inequality: Perceptions of Distributive Justice among Urbanites, Migrants, and Peasants", unpubl. paper, Center for the Study of Democracy University of California, Irvine.

Khan, A. R. (2004), *Growth, Inequality and Poverty in China*, Issues in Employment and Poverty Discussion Paper 15 (Geneva: International Labour Office).

Li, S., Wei, Z., and Jing, S. (2005), *Inequality of Wealth Distribution of Chinese People: An Empirical Analysis of its Cause*, FED Working Paper 65 (Beijing: China Center of Economic Research, Peking University).

Lin, J. Y. (1992), "Rural Reforms and Agricultural Growth in China", *American Economic Review*, 82: 34–51.

Nair, K. R. G. (2007), "Malnourishment among Children in India: A Regional Analysis", *Economic and Political Weekly*, 42(17): 3797–803.

Purfield, C. (2006), *Mind the Gap: Is Economic Growth in India Leaving Some States Behind?*, IMF Working Paper 06/103 (Washington, DC: International Monetary Fund).

Ravallion, M., and Chen, S. (2007a), *Absolute Poverty Measures for the Developing World 1981–2004*, World Bank Policy Research Working Paper 4211 (Washington, DC: World Bank).

—— —— (2007b), "China's (Uneven) Progress against Poverty", *Journal of Development Economics*, 82: 1–42.

Suri, K. C. (2004), "Democracy, Economic Reforms, and Election Results in India", *Economic and Political Weekly*, 39(51): 5404–11.

Topalova, P. (forthcoming), "Trade Liberalization, Poverty and Inequality: Evidence for Indian Districts", in A. Harrison (ed.), *Globalization and Poverty* (Chicago: University of Chicago Press).

World Bank (2006), *World Development Report 2006* (Washington, DC: World Bank).

CHAPTER 19

ECONOMICS, ETHICS AND CLIMATE CHANGE

SIMON DIETZ
CAMERON HEPBURN
NICHOLAS STERN

I. Introduction

> If informed scrutiny by the public is central to any such social evaluation (as I believe is the case), the implicit values have to be made more explicit, rather than being shielded from scrutiny on the spurious ground that they are part of an "already available" metric that society can immediately use without further ado.
>
> A. K. Sen (1999: 80)

There may never have been a kind of environmental pollution to challenge our powers of analysis and evaluation quite like climate change. Global in its causes

We thank Tony Atkinson, John Broome, Simon Caney, Partha Dasgupta, Stephen Gardiner, Roger Guesnerie, Claude Henry, David Newbery, Bill Nordhaus, Tony Venables and Martin Weitzman for discussion of these issues. We are also grateful to participants in the Climate Change and Global Justice conference at the Department of Politics and International Relations, University of Oxford, 21–2 September 2007.

and consequences, spanning many generations, and poorly understood but with the potential to transform the natural environment, climate change has highlighted challenging questions of the economics of risk, of distribution amongst individuals over space and time, and of different aspects of well-being. Any economic analysis of climate-change policy must examine the conflicting interests of different people and, indeed, consider the meaning and relevance of "interest".[1]

It is not enough simply to presume that existing markets can provide a technocratic solution to ethical questions of inter-generational justice. Indeed, by its very nature climate change demands that a number of ethical perspectives be considered, of which standard welfare economics is just one. Other relevant and important approaches highlight rights, freedoms and the prevention of harm, as well as virtues and social contracts.

Furthermore, even if we accept the ethical tenets of the general welfare economics approach, the "workhorse" model of standard welfare economics is but one of several possible specifications. It has the merits of parsimony, but the disadvantages of incorporating a relatively restrictive ethical structure.

Nevertheless, economics can provide valuable guidance on ethical issues by focusing on the consequences of policy and clarifying the implications of particular viewpoints on how to aggregate these consequences. But the relationship between economics and ethics cuts both ways: a careful, explicit examination of the relevant ethical issues can guide the formulation of economic questions on climate change and establish the meaning and relevance of the analysis.

In section II, we set the standard welfare-economic approach to climate-change policy in the context of other major ethical perspectives. In section III, we set out the dimensions of human well-being relevant to most ethical perspectives, along with the value judgements necessary for aggregation across those dimensions. The advantages and disadvantages of the various methods for determining value judgements are assessed in section IV, before these methods are applied in section V to three specific types of aggregation: temporal inequality, intra-temporal inequality and risk. Section VI offers some brief concluding thoughts.

II. Ethical Perspectives

Economists understand climate change to be an externality, just like, for example, noise pollution and traffic congestion. But in comparison with most other externalities, climate change is much more encompassing.

[1] A broader analysis could also give due consideration to the possible interests of non-humans, but this is very distant from familiar approaches in economics, so we simply note as much at this point.

First, it is global in its causes and consequences. With respect to causes, almost all aspects of our daily lives are associated with the emission of greenhouse gases, whether directly or indirectly, and the warming caused by a tonne of greenhouse gases is independent of where in the world it is emitted. Responsibility is thus universal, but not equally shared. There are big differences in the volumes of greenhouse gases emitted in different parts of the world (and also within countries and communities), both in aggregate and per capita. As for consequences, the impacts of climate change are also universal, more obviously so if the worst scenarios are realized. Ecosystems, and their goods and services, depend on climatic conditions, so climate change threatens the basic elements of life for people around the world—access to water, food, health, and the use of land and the environment. But impacts are very uneven. Geographically and economically vulnerable regions—and groups within regions—will experience the worst impacts first. Other regions may even benefit from climate change, if the changes are small and gradual (though that is a big if: see Schellnhuber *et al.* 2006).

Second, the impacts of climate change are long-term and persistent. Once emitted, greenhouse gases can remain in the atmosphere for many decades (carbon dioxide can remain there for one or two centuries), and there is a lag in the response of the climate system to emissions and of ecological, social and economic systems to changes in climate, so the impacts of today's emissions last for centuries, and climatic changes today and in the near future depend on past emissions.

Thus, although ethical considerations are an essential feature of all public policy debates, they are fundamental in a particularly direct and obvious way to climate-change policy. Indeed, there are further ethical questions that climate change raises in common with other environmental externalities, most notably what ethical status should be given to the natural environment.[2] As should already be clear, while not primarily an ethical issue, the great uncertainty about the impacts of climate change and the (lesser) uncertainty about the costs of reducing greenhouse gas emissions are also relevant.[3]

Because of its great reach over many dimensions (e.g. time, space, states of nature and dimensions of human well-being), it is possible to take a wide variety of approaches to the ethics of climate change: standard welfare economics is just one. It is one of many approaches that look only at the consequences of actions for the welfare or "utility" of individuals in a community (perhaps most appropriately

[2] In this paper we will adopt a largely anthropocentric focus. Many important, contemporary contributions to the notion of intrinsic value in nature (that the natural environment is in itself an object of moral consideration) were made in the 1970s by, for example, Naess (1973), Routley (1973), Rolston (1975), Singer (1977) and Goodpaster (1978). For recent reflections on its broader significance, for example to policy, compare Callicot (2002) with Norton (1992). A classic earlier account was written by Leopold (1949), while White's (1967) essay on what he saw to be the damaging environmental ethic of Western philosophy and religion stimulated much of the subsequent work.

[3] See Gardiner (2004) on how uncertainty about climate change interacts with ethical considerations.

described by Sen as "welfarism"), where utility is derived from the consumption of goods and services (assuming utility can be measured by the strength of everyone's preferences and these preferences can be described by the same utility function).

No ethical account is thereby taken of the actions by which consequences are caused. In contrast, "deontological" ethical systems, including concepts of rights and freedoms, focus on the ethical qualities—the "rightness" or "wrongness"—of actions in themselves, often setting out various duties and obligations.[4] Although the consequences of actions play a constitutive role in understanding to what, for example, one has a right or freedom, modern-day proponents emphasize that such approaches can correct the more obviously egregious outcomes of unfettered preference-satisfaction utilitarianism, whereby the satisfaction of aggregated preferences can in principle mask great harm to a minority: for example, "It is quite likely that a cost–benefit analysis in ancient Rome of the spectacle of throwing Christians to the lions in the Colosseum would have come up with a positive result" (Beckerman and Pasek 2001: 4).

The predicted impacts of climate change raise questions of rights and corresponding duties. It is argued that future generations have the right to enjoy a world whose climate has not been transformed to compromise basic physical security (e.g. Shue 1999) or result in other dangers (e.g. Caney 2005, 2006). If this is so, the present generation should reduce greenhouse gas emissions to protect the rights of future generations. In terms of climate change in the next few decades, which largely cannot be avoided, industrialized countries are arguably duty-bound to compensate developing countries for the damage inflicted by past emissions (e.g. Mueller 2006).

Rights to a stable climate might conflict with established rights to, for example, basic education and health, or to common treatment in voting. These established rights have very powerful consequences in terms of law, policy and structures of society. Ideally, all these rights should be fulfilled, but where a conflict of rights arises, one right will be violated and there will be a moral loss. Comparisons of the consequences may help determine which right takes priority. We will return to these issues shortly in discussing how comparisons of the consequences of climate policies are indispensable.

One might also seek to justify emission reductions based on the weaker notion that emitters of greenhouse gases (past, present and future) have obligations—not arising from rights—to consider the climate damage caused, just as a passer-by might be morally obliged to aid someone who has fallen ill, even though the ill person is unlikely to have a right to that assistance as such.[5] Barry (1999)

[4] The distinction between consequentialist and deontological approaches can be overstated. For instance, rule utilitarianism can recognize the moral and social importance of rights and obligations (Harsanyi 1980).

[5] Although in a number of jurisdictions there is in fact a legal requirement for the passer-by to provide assistance, for example in France and Germany.

constructs a theory of inter-generational justice that, he argues, does not depend on equal rights across generations. Instead, it depends only on the twin notions of "responsibility"—that "bad outcomes for which somebody is not responsible provide a *prima facie* case for compensation" (97)—and "vital interests"—namely that there are certain objective requirements that all human beings have, regardless of their location in space or time.

A concept related to the idea of inter-generational obligations is that of sustainable development: future generations should have a standard of living—or at least the opportunities to attain a standard of living—no lower than the current one.[6] Economists have constructed various rules from this definition of sustainable development, in terms of the capital stock passed from one generation to another (Dasgupta and Heal 1974; Solow 1974; Hartwick 1977). These rules broadly focus on maintaining a constant overall stock of all forms of capital, including education, health, capital equipment, buildings, natural resources and the natural environment (some would also include social, cultural or institutional capital). If different forms of capital are substitutable to a degree, it follows that the natural environment need not be perfectly preserved for the next generation if they are provided with compensation in the form of another kind of capital.

Of course, it is impossible for the global environmental and ecological system, which provides us with life-support functions such as stable and tolerable climatic conditions, to be replaced in its entirety. This has led to the development of special rules for the preservation of essential or so-called "critical" environmental assets. The commitment of Article 2 of the United Nations Framework Convention on Climate Change to "achieve stabilisation of greenhouse gas concentrations at a level that would prevent dangerous anthropogenic interference with the climate system" can be interpreted as just such a sustainability rule. The challenge of course is to establish what environmental assets are "critical".

The notion that moral obligations may exist without corresponding rights is also inherent in "virtue ethics", which emphasizes virtuous character as a guide to moral behavior, rather than focusing on consequences or rules. These approaches have enjoyed a renaissance in the past half-century of Western moral philosophy, in large part because they may offer a more plausible explanation of how individuals actually think about ethics (Anscombe 1958; Foot 1978; MacIntyre 1981; Wiggins 2006). In the context of climate change, Jamieson (1992) has used virtue ethics to argue that our existing values are not up to the task presented by climate change, and that a change in values—virtues—is the only promising ethical approach. Interestingly, Crisp (1992) argues that utilitarianism (a consequentialist ethic) itself recommends pursuing a "life of virtue", and in the specific case of climate change, Jamieson (2007) insists that "utilitarians should be virtue theorists".

[6] Jamieson (1998) provides a philosophical critique of the notion of sustainability.

Irrespective of which broader ethical framework is adopted, all frameworks will need some type of comparison of the consequences of climate change policies. For instance, we are likely to have to face trade-offs between minimal rights today to a basic standard of living, which might plausibly be compromised by very strong, very rapid reductions in greenhouse gas emissions, and counterpart rights in the far-off future threatened by climate change. Further, everyday or "first-order" morality (Wiggins 2006) would seem to include evaluating trade-offs in a consequentialist manner. As Sen noted, "the general case for taking full note of results in judging policies and institutions is a momentous and plausible requirement" (1999: 61).

III. Dimensions of Human Well-Being

Different notions of consequentialist ethics emphasize different aspects of the consequences of decisions for present and future people and for the environment. Sen (1999) argues persuasively that relevant consequences not only include those for human well-being, but also those for human *agency*: individuals may seek to maximize objectives beyond their own well-being. In particular, Sen's focus on human capabilities, as well as outcomes, is relevant to climate-change policy and economics. While we do not pursue the capabilities approach here, we nevertheless consider it to have substantial potential as an ethical framework for climate-change policy evaluation.

The lists of dimensions of human well-being that most consequentialist approaches would incorporate have strong similarities, above all on the four dimensions of consumption, education, health and the environment. These are usually the areas of focus in cross-country comparisons of living standards, such as, for example, the World Development Indicators of the World Bank, the Human Development Index of the UNDP, and the UN Millennium Development Goals. Through these choices of data and corresponding goals, the international community has identified a strong and shared view of some of the key dimensions of human well-being.

How the implications on these four dimensions are assessed and aggregated, will, of course, vary according to the ethical position adopted. In making assessments, how and whether we attempt to aggregate further over consequences (1) within generations, (2) over time and (3) according to risk will be crucial to policy design and choice. When we do aggregate explicitly, we have to compare quantitatively consequences of different kinds and for different people. These comparisons pose a new set of questions and problems.

While aggregation over the relevant dimensions helps to develop an understanding of the relative merits of different climate-change policies, it is also clear that many of the important issues arising from climate change are most obvious prior to aggregation into a single index such as world consumption. The aggregation will usually have to ignore or suppress crucial information, since data on key consequences or variables are unlikely to be available in a form amenable to aggregation for all dimensions.

Many decision-makers may feel able to formulate policies without an attempt at formal modeling, which explicitly aggregates effects over time, space, people and possible random outcomes. The Stern Review (Stern 2007) argued that decision-makers could form a reasoned view on whether they would spend (say) 1% of income a year to avoid the risk of the worst consequences of climate change. They can form that view from a description of these possible consequences, without necessarily attaching numbers to all the consequences, weighting them and adding them all up. In this more disaggregated approach, decision-makers simply take a view on the broad description of future effects and do not need, for example, to specify social discount rates.

There will be some kind of implicit aggregation in many, but far from all, of the relevant judgements or decisions (for example, for some decisions only partial orderings may be needed, with very limited or no aggregation). But bringing aggregation out formally in terms of values, where it is otherwise implicit, can be helpful in highlighting areas of agreement and disagreement, as demonstrated by the discussion since the publication of the Stern Review (see Dietz, Anderson *et al.* 2007 and Dietz, Hope, Stern *et al.* 2007; Beckerman and Hepburn 2007).

IV. Methods for Determining Values

When decision-makers do adopt an explicit aggregation process, an ethical framework must be selected and the values for explicit ethical parameters must be determined. Understandably, moral and political philosophers have had a great deal more to say about the derivation of moral values than economists.[7] Most frequently, economists simply adopt a consequentialist, and specifically welfarist, perspective, where a social welfare function that is additive over individual utilities (without any weights) is specified. Then, in the case where a narrow consumption-based

[7] An inadequate and highly incomplete list would include the "reflective equilibrium" notion of Rawls (1971), the "argument from received opinion" discussed by Hare (1971), the use of opinion polls (Miller 1999), the theory of "discourse ethics" (Habermas 1990) and works by Barry (1995), Griffin (1996) and others.

approach is followed, one assumes a particular form of utility function (such as the isoelastic function). Finally, within that particular specification, particular parameters need to be determined, such as the pure time discount rate δ and the social elasticity of marginal utility of consumption (or income) η. Yet each of these particular decisions has significant ethical implications. And an important metaethical question arises in making these explicit ethical choices: whose values should determine the ethical framework and the ethical parameters?

IV.1 Whose Value Judgements?

The answer might be thought to depend, at least in part, upon the character of value judgements. One might start, for instance, from the position that certain ethical statements have an "objective" character, so that careful thought ultimately leaves "nothing else to think" (Wiggins 1987, 2006).[8] If this is so, then the views of expert moral philosophers—who have devoted more time and careful consideration to nuanced ethical questions than most of us—might be accorded particular weight.[9] On the other hand, even the most distinguished philosophers disagree on important ethical questions—Aristotle's defence of slavery in the *Politics* provides a case in point.

However, the fact that general ethical statements may be answered differently by different societies does not necessarily imply subjectivism (Mackie 1977). Wiggins (2006: 347) draws an important and nuanced distinction between relativism and contextualism, pointing out that "no act or practice can be assessed as right or wrong, good or bad, etc. without the full specification of circumstances and context". For instance, humans might universally agree that immoral practices in one society are nevertheless perfectly moral in another. In other words, it is important to distinguish between *general* moral claims (in contrast to *context-specific* moral claims) and *universal* moral claims (which are true from all vantage points). Moral objectivism, according to Wiggins (2006), amounts to the modestly optimistic claim that fully contextualized ethical propositions may find universal support.

Where ethical agreement cannot be found, some mechanism is needed to mediate between different perspectives and aggregate different value judgements for use in public decision-making. Should the views of experts be accorded greater weight than those of the general public? There may be a parallel with the economic literature on biodiversity valuation, where Nunes and van den Bergh (2001) note the

[8] For instance, in arguing for the fundamental nature of "equality", but in stressing the question "equality of *what*?", Sen (1992: 18–19) notes that "impartiality and equal concern, in some form or other, provide a shared background to all the major ethical and political proposals in this field that continue to receive argued support and reasoned defence".

[9] Even if moral philosophers are the best judges of ethical questions, political philosophers would note that it does not follow from the proposition that "X is the best judge" that "X should make the decision (or have a greater say)".

argument that lay people cannot judge the relevance and complexity of biodiversity and ecosystem functioning, so valuation should be left to experts such as biologists. In contrast, Arrow *et al.* (1993) recommend that lay people be sufficiently informed by experts before they express their valuations. Similarly, one might recommend that we should aggregate the ethical views of the general public, suitably informed and supported by expert philosophical guidance.

IV.2 Stated or Revealed Ethics?

Ethical judgements and parameters might be determined through "stated" approaches, which rely on people's explicit responses to questions or hypothetical thought experiments, or "revealed" approaches, which attempt to deduce implicit ethical values from observed behavior. Economists are familiar with the notion that behavior can reveal implicit individual "preferences", and some economists would further argue that appropriate ethical judgements can be deduced from these revealed preferences.

To deduce ethical values from preferences revealed by behavior, at least four (non-trivial) conditions would be required:

1. a unique preference is revealed by the observed behavior;
2. the preferences revealed are the "true preferences" of the individual, based on full and correct information without any errors in decision-making;
3. the preferences measured are contextually relevant to the ethical judgement at hand; and
4. the preferences are appropriate for social decision-making, and not merely individual decision-making.

If these four conditions are satisfied, the behavior might reveal an appropriate ethical view. The merit of the "revealed ethics" approach is that the resulting ethical judgements are based upon real decisions, rather than hypothetical speculation. However, the four conditions set out above are rarely likely to be satisfied.

Condition (1) is unlikely to be satisfied because "revealed ethics" approaches suffer from the "inverse optimum" problem: underlying ethical values must be inferred from the observed behavior, much as in standard consumer demand theory and estimation, and several different ethical perspectives may be consistent with the same behavior. The outcome of this process depends on assumptions about model structure and the representation and parameterization of values. Stated approaches also suffer from an "inverse optimum" problem, but it is not as severe, for the simple reason that stated approaches can be designed to elicit responses to the direct ethical issue at hand.

The problem with satisfying condition (2), as Harsanyi noted, is that revealed preferences, or "manifest preferences", may be "based on erroneous factual beliefs,

or on careless logical analysis, or on strong emotions that at the moment greatly hinder rational choice". In contrast, "true preferences" are the preferences that an individual would have if "he had all the relevant factual information, always reasoned with the greatest possible care, and were in a state of mind most conducive to rational choice" (Harsanyi 1982: 55).[10] The distinction is not merely theoretical: people consistently mispredict the factors that determine their quality of life (see e.g. Kahneman, Wakker and Sarin 1997), and behave in a wide variety of suboptimal and irrational ways documented by psychologists and behavioral economists. In a non-welfarist and non-utilitarian calculus, there may be merit in accepting people's errors and accounting for people's actual preferences, warts and all, because of the intrinsic value of the autonomy to make our own decisions. Interestingly, Harsanyi (1982: 56) went on to argue that "manifest preferences" should count in assessments of someone's welfare as "our final criterion in judging what his real interests are and what is really good for him". It is, however, perhaps more questionable whether imperfect information and errors of analysis should be incorporated into an exercise intended to reveal ethical judgements.

Just as context matters to the formation of ethical judgements, as discussed above, it is also crucially important to observed behavior, which makes condition (3) unlikely to be satisfied. Levitt and List (2007: 162) conclude that behavior is "affected by a dazzlingly complex set of relational situations, social norms, frames, past experiences, and the lessons gleaned from those experiences". If the objective is to determine an appropriate set of ethical judgements to inform climate-change policy, the revealed preferences should be derived from individual behavior in a context that maps appropriately onto an international, inter-generational, non-marginal, risky and uncertain environmental policy problem. Behavior revealed in one setting does not necessarily map neatly onto other settings, just as insights from a carefully controlled scientific laboratory experiment in a vacuum do not apply perfectly to a real-world situation with friction. And human behavior is even less likely to be replicated outside the laboratory (Levitt and List 2007).

Fourth, Sen (1992: 19) clearly articulates the distinction between questions of political philosophy regarding social arrangements, on the one hand, and those of personal ethics on the other, putting condition (4) into question. He notes that in "the ethics of *personal* behaviour, powerful arguments have been presented in favour of permitting or requiring explicit *asymmetries* in the treatment of different people. Such arguments may relate, for example, to the permissibility—perhaps even the necessity—of paying special attention to one's own interests . . . [and] . . . one's own family members and others to whom one is 'tied' ". It is not impossible, for instance, that many people behave according to some agent-relative ethic in their individual decisions, placing a greater weight on the interests of family, friends and

[10] Sen and Williams (1982) also note that individuals' actual preferences may, deliberately or accidentally, be very different from the preferences that would maximize their well-being.

others within close communities. And yet, in considering what "we" should do about policy issues such as climate change, the same people might wish to adopt an impartial ethic. More generally, the ethics revealed from individual behavior may be an inappropriate guide to the same individual's ethical values for social decisions.

For these four reasons, great caution is warranted when using "revealed ethics" approaches to determine appropriate value judgements for social decision-making.[11] This is not to ignore the problems in "stated ethics" approaches, which mirror the problems documented in the contingent valuation literature.[12] But "stated ethics" approaches are less likely than "revealed ethics" approaches: (1) to face difficulty from the inverse optimum problem; (2) to misrepresent true preferences; (3) to be contextually inappropriate; and (4) to reveal personal ethics rather than values for political philosophy. Stated approaches also provide more opportunity than revealed approaches for philosophical expertise to inform ethical judgements.

Whether a stated or revealed ethics approach is employed, a mechanism is required to aggregate the views of the individuals whose preferences are deemed to count. Economists reach almost instinctively for the markets when in need of an aggregation device, so we now consider the relevance of market data.

IV.3 The Relevance of Market Data

Markets (at least in the narrow sense of aggregate market demand and supplies) provide a "revealed ethics" approach that automatically aggregates individual preferences. They may therefore appear to be an obvious device to aggregate individual value judgements about social problems such as climate change. Indeed, some economists have asserted that deep ethical questions involved in climate-change policy, such as the appropriate weight to place on the interests of future generations, should be resolved by reference to market data (Nordhaus 2007; Weitzman 2007). However, there are at least five additional reasons, beyond the four caveats discussed above, to be cautious in employing market data for ethical values generally, and for climate-change ethics in particular.

First, the standard problems of market imperfections distort market valuations via shadow prices or social valuations. Drèze and Stern (1987, 1990) provide a detailed and formal analysis of relations between market imperfections, shadow prices and market prices in a general equilibrium framework.

Second, markets aggregate preferences with a weighting applied to those who participate more heavily. This tends to imply that the preferences of the wealthy

[11] Harsanyi (1982: 56) might be expressing a fifth when he says that "we must exclude all clearly antisocial preferences, such as sadism, envy, resentment, and malice".

[12] See e.g. Carson (2000).

(rather than the more "expert", for instance) count more in determining overall social ethical judgements. For example, actual savings rates do not represent equally aggregated ethical judgements in an optimization problem, but instead are the result of a complex process of interaction, where different agents have different endowments and bargaining power, and where the decisions of those with more wealth have a greater impact on overall savings rates.

Third, futures markets exist only over a few decades and do not cover the very long periods, i.e. centuries, relevant for the discussion of climate change. Markets therefore cannot directly reveal appropriate ethical parameters beyond their horizons. Furthermore, as markets only aggregate the preferences of the people participating in them, the interests of future generations are not incorporated, even though they are the ones most profoundly affected by climate change.

Fourth, and this is related to the fourth general condition about "revealed ethics" in section IV.2, individual decision-making in markets will (appropriately) reflect individual rather than social concerns. Market data do not reflect an answer to the question of what citizens of a society should do when considering together what they would regard as the right or responsible action. For instance, the probability of demise for an individual is much higher than that of the probability of demise for humanity as a whole (which is relevant to the appropriate pure time discount rate, δ).

Fifth, markets are unable to capture the ethical issues associated with non-marginal and irreversible change at the global level, since any one individual's action will not affect the set of aggregate circumstances.

In our view, the issues we face as a result of climate change raise very difficult philosophical questions, versions of which have challenged philosophers across the ages. There is no alternative to engaging in a normative discussion. Some may seek to argue the case that markets are the appropriate device both to reveal and aggregate ethical judgements. Many others will find this an implausible position. Irrespective of that matter, the ethical issues cannot be side-stepped, even by those wanting to resort to market information to reveal the answers to philosophical questions.

V. Values for Aggregation

We now illustrate some of the above issues using the very simple structures from welfare economics as it has come to be standardly applied to climate change in the past 20 years or so (Cline 1992; Nordhaus 1994; Manne *et al.* 1995; Tol 1997; and others).

V.1 Model Structure

In brief, the standard, "workhorse" model of welfare economics in this case specifies a particular Bergson–Samuelson social welfare function that is additive over the utilities of all individuals in all time periods (without any weights):

$$W = \sum_{t=0}^{T} \sum_{i=1}^{N_t} (u_{i,t})(1+\delta)^{-t}, \tag{1}$$

where W is aggregate social welfare, u is the utility of individual i, and there are N_t individuals in time period t. It is simplest in this case to think of the specification for the function $u(.)$ as a social representation and not a cardinal measure of utility. Importantly, the contribution of utility to social welfare in different time periods is a function of the utility discount rate or pure rate of time preference, δ. This is a weighting of utilities in different periods purely because of the different dates and is quite separate from the circumstances of individuals at these times. In the Stern Review (Stern 2007), this was set to 0.1% per annum. Cline (1992) similarly set δ equal to 0. However, doing so has revealed a notable disagreement between climate-change economists, many of whom (e.g. Nordhaus 2007; Weitzman 2007) prefer to set δ in the region of 2–3% per annum. We shall outline the nature of this disagreement below.

To simplify the analysis greatly, the utility of individuals in major regions of the world is aggregated into a single, representative individual and multiplied by the size of that region's population:

$$W = \sum_{t=0}^{T} \sum_{j=1}^{J} (u_{j,t} N_{j,t})(1+\delta)^{-t}, \tag{2}$$

where j is the region, of which there are a total of J. Note that this formulation depends on the assumption that the population is exogenous. If the total size of the population may be measurably affected by climate change and in turn by policy, more difficult ethical issues arise (see Dasgupta 2001; Broome 2004, 2005).

The utility of the representative individual in each region is described by a utility function, where utility depends only on the aggregate consumption of goods and services via an isoelastic function. The impacts of climate change are expressed relative to a loss in this consumption:

$$u_{j,t} = u(c_{j,t}) = \frac{c_{j,t}^{1-\eta}}{1-\eta}, \tag{3}$$

where c is the consumption of the representative individual for region j. The assumptions behind the workhorse model imply that the elasticity of marginal utility with respect to consumption, η, is effectively a coefficient of social inequality

aversion, in particular to the distribution of consumption across individuals. In a model with many countries and many time periods, inequality is relevant across space at a given point in time (intra-generational distribution) and over time (intertemporal allocation). Further, under uncertainty about future consumption that can be described by a distribution of subjective probabilities, η is a coefficient of relative risk aversion:

$$\mathrm{E}(W) = \sum_{s=1}^{S} \sum_{t=0}^{T} \sum_{j=1}^{J} (u_{j,t,s} N_{j,t,s})(1+\delta)^{-t} p_s, \qquad (4)$$

where s is a particular climate-change scenario (and there are S scenarios), the probability of each scenario being p_s, and $\mathrm{E}(W)$ denotes the fact that the weighted sum of all utilities is now a measure of expected utility.

So, because of the simplifications employed in the conventional model, the value for η captures our attitudes to risk, inequality within generations, and inequality between generations. In the base case of the Stern Review and indeed in many other welfare-economic analyses of climate change, η is set to unity (which gives a special form for equation (3), namely $u_j = \ln c_j$). Any particular choice of η is likely to be contested, including $\eta = 1$, and thus sensitivity analysis is important. Part of the problem in interpreting and discussing η arises because the simplifying assumptions of the standard model leave η to reflect three related but distinct concepts simultaneously. This is problematic, because a great deal of information appears to be of relevance to η and yet it is contradictory.

Indeed, it is not intuitively obvious whether an increase in η would produce an increase or a decrease in the calculated impact of climate change on social welfare. First, higher η implies greater risk aversion, which in turn implies greater concern about the worst climate-change scenarios, and a higher impact of climate change on social welfare. Second, higher η also implies greater social inequality aversion, increasing the relative weight placed on changes in the consumption of the poor, so that greater weight is placed on climate impacts in the poorest regions. Since climate change is expected to hit the poorest regions hardest in relative terms, higher η in this respect increases the overall loss in social welfare. Third, if we also assume, as is standard, that aggregate consumption will continue to grow despite climate change, then the overall loss in social welfare falls with higher η, because it reduces the weight placed on future consumption, and increases the weight placed on present consumption (because the present is poorer than the future). This effect reduces the estimated net impact of climate change, because the impacts of climate change will be mainly felt in the (relatively rich) future. Dietz, Hope and Patmore (2007) model the contrasting effects of risk aversion and intertemporal inequality aversion.

Is there any justification for using the one parameter to reflect aversion to risk, aversion to intra-temporal inequality, and aversion to inequality through time?

Harsanyi (1953) thought so. He considered the perspective of a Kantian impartial observer (or someone in the "original position" (Rawls 1971)), with an equal chance of being born into the place of any individual. He argued that this suggested "a close affinity" between individual risk preferences and social inequality preferences. The analogy between risk and inequality from behind the "veil of ignorance" has proven powerful (Sen 1978; Broome 1990). However, it is a substantial leap to conclude that the two concepts are identical. Empirical evidence suggests the concepts are distinct (e.g. Carlsson *et al.* 2005). At an individual level, Beckerman and Hepburn (2007) note that a person might not be very risk-averse (low η) yet might be strongly averse to inequality (high η). And a risk-averse person (high η) may care little about inequality of outcomes (low η). At a social level, we might be very averse to climate risks (high η), while accepting inequalities in outcomes (low η) because, say, we also value equality of opportunities or capabilities.[13]

Research on subjective well-being further suggests an important difference between inequality within a generation, and inequality between generations. Di Tella and MacCulloch (2006) conclude that evidence on happiness is consistent with the notion that happiness is derived from relative levels of consumption, suggesting that spatial or intra-generational inequality matters. But the evidence also reveals that aggregate happiness remains relatively constant over time (Easterlin 1974, 1995), suggesting that inequality in consumption over time is less important.

These reasons suggest that combining value judgements on risk, intertemporal inequality and intra-temporal inequality into one parameter may be a non-trivial simplification. It is certainly not necessary. Kreps and Porteus (1978), Selden (1978) and Epstein and Zin (1989) specify models that disentangle risk aversion from intertemporal substitution and these models may be better suited to economic analysis of climate change. This is an area for further research. With this in mind, we now examine arguments for different values of η derived from considerations about risk, space and time respectively.

V.2 Risk

How can we use observed attitudes to risk to illuminate the choice of η? Gollier (2006), in reviewing the Stern Review's welfare-economic analysis, has argued that some studies of insurance point to higher values of η, from 2 to 4. Equally, however, some interpretations of gambling behavior suggest low or negative η. And much

[13] Harsanyi's (1955) theorem shows that if social preferences satisfy the Pareto principle for prospects (which amounts to weak separability across people), and they satisfy expected utility theory (which amounts to strong separability across states of nature), then risk aversion is the same as inequality aversion. For people's preferences to embody a degree of risk aversion that is different from their degree of inequality aversion, their preferences either do not satisfy expected utility theory, or they do not satisfy the Pareto principle for prospects.

theoretical and empirical work on choice and risk has argued that the expected-utility model performs very badly indeed. It is very difficult to use empirical evidence on risk aversion to narrow down a relevant range for η. The reasons are: (1) the expected-utility model so often contradicts the data; (2) where it does not contradict the data, the range of possible values of η is very wide; and (3) as discussed in section IV, the coefficients of risk aversion, as revealed by aggregated market decisions, may have limited normative significance for the economics of climate change.

V.3 Intra-Temporal Allocation

Dasgupta (2006) argues on normative grounds for values of η from 2 to 4. In contrast, Atkinson and Brandolini (2007) point out that, if one examines the behavior of public decision-making, there are plausible reasons to think that η is lower than unity in practice. Following Okun's (1975) "leaky bucket" experiment, consider the desirability of a transfer from individual A to individual B, where A has income, say, twice that of B. If $\eta = 1$, so that the elasticity of social marginal utility of income falls as the reciprocal of income (using income and consumption interchangeably, for simplicity), then a transfer of a unit of income from A to B increases welfare provided not more than 50% is lost "along the way" or "leaked out of the bucket". A recent estimate of income distribution for the USA suggests that income at the 90th percentile (i.e. the income of a person earning more than 90% of the rest of the population) is roughly twice the income at the 50th percentile, which is again roughly twice that at the 25th percentile. Assuming $\eta = 1$, a transfer from the 90th to the 10th percentile would be acceptable provided no more than 87.5% is "lost along the way". The "loss along the way" can stand for many possibilities, such as the actual administrative costs of making transfers, disincentives, corruption and so on. Those who consider a loss of 87.5% to be too high would consider 1 to be too high for η in this context.[14]

Further, Atkinson and Brandolini (2007) argue that assuming η to be constant with changing consumption or income is unhelpfully restrictive. For instance, the appropriate elasticity may be low or zero (constant marginal utility of income if zero) for people living below the poverty line. And many would argue that there is very little social benefit in redistributing income amongst the rich (e.g. within the top 5% of the distribution). These views would point to an elasticity that first rises as income rises, then falls. We would not necessarily insist on a curve of this shape, but we would more generally highlight the straitjacket of a constant η. The primary argument for employing constant η is convenience, but we must be aware of its limitations. We should also note that in many examples of cost–benefit analysis of

[14] If the commonly used Gini coefficient is interpreted as a welfare index (e.g. Sen 1997), the appropriate curve would be much flatter, in this sense similar in effect to η lower than unity.

projects as practiced by public-sector economists, there is no weighting for intra-temporal income differences: in other words η is equal to zero.

Our conclusion from this discussion of intra-generational transfers and inequality measures is that, if a constant elasticity of consumption measure is to be used, then while some observers would view 1 to be too low for η, others would consider it rather high. On the other hand, with a non-constant η, the value could be higher than 1 in some parts of the income range and lower than 1 elsewhere.

V.4 Time

Building on the model structure above, many find it helpful to think about inter-temporal value judgements in terms of the social rate of time preference or social discount rate. The social discount rate r in this model is:

$$r = \eta g + \delta, \tag{5}$$

where g is the rate of consumption growth per capita. We must clearly distinguish between r, the social discount rate, and δ, the pure time or utility discount rate. The latter is an integral feature of the canonical model of social welfare presented above and is a primary ethical object. That is, its ethical status requires careful consideration. The former is the discount rate on consumption, not utility, and can be used where the project under evaluation is marginal to the path of future consumption. We cannot make this assumption in the case of climate change, where the impacts are potentially *non-marginal* in that they could significantly reduce the growth rate (inclusive of the impacts of climate change on income and welfare) at some point in the future.

As noted, δ has a direct ethical meaning: it determines the weight placed on the utility of people born at different times. The classical utilitarian approach in equation (1), taking the sum of (unweighted) utilities, assumes that individuals should be treated equally and symmetrically. Discrimination by date of birth is then difficult to justify. Yet if $\delta = 3\%$ per annum, as suggested by some climate-change economists (e.g. Nordhaus 2007), someone born in 1960 should "count" for roughly twice someone born in 1985. This is implausible. The Stern Review (Stern 2007) argued that the most straightforward ethical approach to understanding δ was through the interpretation of $(1 + \delta)^{-t}$ as the probability of existence of future human civilizations, and δ was set to 0.1% per annum. Many would see the implied 90% chance of human civilization seeing out this century as alarmingly low, although Rees (2003) has put the chance at 50%.

However, the combination of an ethical choice such as $\delta = 0.1\%$ with an elasticity of marginal utility of $\eta = 1$ places quite heavy weight on benefits far into the future. One way of illustrating this is to note that if the long-run growth rate of consumption is g (and suppose that eventually population growth is zero), then

the utility integral over infinity converges if and only if

$$\delta > (1 - \eta)g \tag{6}$$

(see Stern 2007: 58, 670). If η is less than 1 and δ is low, this condition will be violated.[15] Cline (2007), in his review of the Stern Review, has emphasized that the combination of $\eta = 1$ and $\delta = 0.1\%$ implies that the weight placed on the impacts of climate change in the very far-off future is large. The empirical models used by climate-change economists typically have a fixed time-horizon of one to three centuries,[16] so the impacts in question are extrapolations from beyond the end of the modeling horizon. This would be less of a problem if δ and η were higher, because the share of the present value of climate impacts in the distant future would be lower (Dietz, Anderson *et al.* 2007). However, this is not an appropriate reason for the choice of ethical parameters.

If an ethical observer knows what the future ethical savings rate should be, then there is another link between η and δ. Dasgupta (2006) and Nordhaus (2007) compare the optimal savings rate implied by $\eta = 1$ and $\delta = 0.1\%$ (as in the Stern Review) with observed savings rates today. They suggest that the implied optimal savings rates are higher than observed rates, and conclude that this combination of parameters is inappropriate. There are three responses. First, currently observed savings rates may not be long-run ethical savings rates. Second, there are many savings rates to observe, from around zero or below in the USA, to over 50% in China. Third, this is an "inverse optimum problem", where the conclusion is dependent on the model structure assumed. In a very simple model of production where consumption growth derives solely from the returns to investment in (man-made) capital, $\eta = 1$ and $\delta = 0.1\%$ would lead to very high optimal savings rates. But adding in technical progress can bring optimal savings rates down to current levels in industrialized countries. This was known to Keynes and Ramsey and is considered in some detail by Mirrlees and Stern (1972).

VI. Conclusion

Climate change has highlighted challenging questions in the economics of risk, of distribution amongst individuals, of distribution over time and generations, and of different aspects of well-being. Explicit aggregation of all effects into one model

15 In assessing this observation we should, however, note that the condition for convergence applies to the very long run or to limiting values of the elements in equation (6). There is no reason to suppose that any of them are constant in the short, medium or long run.

16 For example, the PAGE model used by the Stern Review runs until 2200 (Hope 2006).

is not necessary for a reasoned judgement of whether the costs of taking action on climate change justify the benefits of damages avoided. However, whichever approach is followed, reasoned judgement is impossible without considering difficult ethical issues, taking most economists into unfamiliar territory. We are fortunate to have intellectual leaders such as Amartya Sen to provide us with guidance on the boundaries of ethics and economics.

This guidance reveals several important insights. Because climate change is a non-marginal, international, inter-generational policy challenge involving important questions of risk and uncertainty, our usual shortcuts are inappropriate. Moreover, existing markets do not provide a technocratic solution to questions of intergenerational justice. Instead, climate change demands that a number of ethical perspectives be considered. Standard welfare economics is one such perspective. Other perspectives, not least Amartya Sen's capabilities approach, deserve more detailed application beyond the ambit of this paper.

Even if we accept the ethical tenets of the general welfare economics approach, this paper has argued that the "workhorse" model of standard welfare economics has a relatively restrictive structure. The simplifying assumptions that make the model so tractable and useful also imply that three distinct and important ethical concepts are represented in one parameter. Specifying a richer welfare-economic model is likely to have a significant impact on results in the economics of climate change.

Economics can provide valuable guidance on ethical issues by focusing on the consequences of policy and clarifying the implications of particular viewpoints on how to aggregate these consequences. In this respect, sensitivity analyses in formal models perform the very important function of allowing the consequences of different ethical parameters to be explored. But the relationship between economics and ethics cuts both ways: a careful, explicit examination of the relevant ethical issues can guide the formulation of economic questions on climate change and illuminate the meaning and relevance of the analysis.

References

Anscombe, G. E. M. (1958), "Modern Moral Philosophy", *Philosophy*, 33(124): 1–19.

Arrow, K. A., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H. (1993), *Report of the NOAA Panel on Contingent Valuation*, available at <www.darrp.noaa/gov/library/pdf.cvblue.pdf>.

Atkinson, T., and Brandolini, A. (2007), "On Analysing the World Distribution of Income", unpubl. paper, Oxford University.

Barry, B. (1995), *Justice as Impartiality* (Oxford: Oxford University Press).

—— (1999), "Sustainability and Intergenerational Justice", in A. Dobson (ed.), *Fairness and Futurity* (Oxford: Oxford University Press).

Beckerman, W., and Hepburn, C. J. (2007), "Ethics of the Discount Rate in the Stern Review", *World Economics*, 8: 187–210.

—— and Pasek, J. (2001), *Justice, Posterity, and the Environment* (Oxford: Oxford University Press).

Broome, J. (1990), *Weighing Goods* (Oxford: Basil Blackwell).

—— (2004), *Weighing Lives* (Oxford: Oxford University Press).

—— (2005), "Should We Value Population?", *Journal of Political Philosophy*, 13: 399–413.

Caney, S. (2005), "Cosmopolitan Justice, Responsibility and Global Climate Change", *Leiden Journal of International Law*, 18(4): 747–75.

—— (2006), "Global Justice, Rights and Climate Change", *Canadian Journal of Law and Jurisprudence*, 19: 255–78.

Carlsson, F., Daruvala, D., and Johansson-Stenman, O. (2005), "Are People Inequality-Averse, or just Risk-Averse?", *Economica*, 72(287): 375–96.

Carson, R. T. (2000), "Contingent Valuation: A User's Guide", *Environmental Science and Technology*, 34: 1413–18.

Cline, W. (1992), *The Economics of Global Warming* (Washington, DC: Institute for International Economics).

—— (2007), "Evaluating the Stern Review on the Economics of Climate Change", unpubl. paper, Center for Global Development and Peterson Institute for International Economics.

Crisp, R. (1992), "Utilitarianism and the Life of Virtue", *Philosophical Quarterly*, 42: 139–60.

Dasgupta, P. (2001), *Human Well-Being and the Natural Environment* (Oxford: Oxford University Press).

—— (2006), "Comments on the Stern Review's Economics of Climate Change", unpubl. paper, University of Cambridge.

—— and Heal, M. G. (1974), "The Optimal Depletion of Exhaustible Resources", *Review of Economic Studies*, 41, suppl.: 3–28.

Dietz, S., Anderson, D., Stern, N., Taylor, C., and Zenghelis, D. (2007), "Right for the Right Reasons: A Final Rejoinder on the Stern Review", *World Economics*, 8: 229–58.

—— Hope, C., and Patmore, N. (2007), "Some Economics of 'Dangerous' Climate Change: Reflections on the Stern Review", *Global Environmental Change*, 17: 311–25.

—— —— Stern, N., and Zenghelis, D. (2007), "Reflections on the Stern Review (1): A Robust Case for Strong Action to Reduce the Risks of Climate Change", *World Economics*, 8: 121–68.

Di Tella, R., and MacCulloch, R. (2006), "Some Uses of Happiness Data in Economics", *Journal of Economic Perspectives*, 20: 25–46.

Easterlin, R. A. (1974), "Does Economic Growth Improve the Human Lot? Some Empirical Evidence", in P. A. David and M. W. Reder (eds), *Essays in Honor of Moses Abramovitz* (New York: Academic Press).

—— (1995), "Will Raising the Incomes of All Increase the Happiness of All?", *Journal of Economic Behavior and Organization*, 27: 35–47.

Epstein, L. G., and Zin, S. E. (1989), "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", *Econometrica*, 57: 937–69.

Foot, P. (1978), *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Blackwell).

Friedman, M., and Savage, L. J. (1948), "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, 56: 279–304.

Gardiner, S. (2004), “Ethics and Global Climate Change”, *Ethics*, 114: 555–600.
Gollier, C. (2006), “Institute Outlook: Climate Change and Insurance. An Evaluation of the Stern Report on the Economics of Climate Change”, unpubl. paper, Barbon Institute.
Goodpaster, K. E. (1978), “On Being Morally Considerable”, *Journal of Philosophy*, 75: 308–25.
Griffin, J. (1996), *Value Judgement: Improving our Ethical Beliefs* (Oxford: Clarendon Press).
Habermas, J. (1990), *Moral Consciousness and Communicative Action* (Cambridge, Mass.: MIT Press).
Hare, R. M. (1971), *Essays on Philosophical Method* (Berkeley, Calif.: University of California Press).
Harsanyi, J. C. (1953), “Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking”, *Journal of Political Economy*, 61: 434–5.
—— (1955), “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility”, *Journal of Political Economy*, 63: 309–21.
—— (1980), “Rule Utilitarianism, Rights, Obligations and the Theory of Rational Behavior”, *Theory and Decision*, 12: 115–33.
—— (1982), “Morality and the Theory of Rational Behavior”, in A. K. Sen and B. Williams (eds), *Utilitarianism and Beyond* (Cambridge: Cambridge University Press).
Hartwick, J. M. (1977), “Intergenerational Equity and the Investing of Rents of Exhaustible Resources”, *American Economic Review*, 67: 972–4.
Hope, C. (2006), “The Marginal Impact of CO_2 from PAGE2002: An Integrated Assessment Model Incorporating the IPCC’s Five Reasons for Concern”, *Integrated Assessment*, 6: 19–56.
Jamieson, D. (1992), “Ethics, Public Policy and Global Warming”, *Science, Technology and Human Values*, 17: 139–53.
—— (1998), “Sustainability and Beyond”, *Ecological Economics*, 24(2–3): 183–92.
—— (2007), “When Utilitarians Should be Virtue Theorists”, *Utilitas*, 19: 160–83.
Kahneman, D., Wakker, P. P., and Sarin, R. (1997), “Back to Bentham? Explorations of Experienced Utility”, *Quarterly Journal of Economics*, 112: 375–405.
Kreps, D. M., and Porteus, E. L. (1978), “Temporal Resolution of Uncertainty and Dynamic Choice Theory”, *Econometrica*, 46: 185–200.
Leopold, A. (1949), *A Sand County Almanac, and Sketches Here and There* (New York: Oxford University Press).
Levitt, S. D., and List, J. (2007), “What do Laboratory Experiments Tell us about the Real World?”, *Journal of Economic Perspectives*, 21: 153–74.
MacIntyre, A. (1981), *After Virtue: A Study in Moral Theory* (London: Duckworth).
Mackie, J. L. (1977), *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin).
Manne, A. S., Mendelsohn, R., and Richels, R. G. (1995), “MERGE: A Model for Evaluating Regional and Global Effects of GHG Reduction Policies”, *Energy Policy*, 23: 17–34.
Miller, D. (1999), *Principles of Social Justice* (Cambridge, Mass.: Harvard University Press).
Mirrlees, J. A., and Stern, N. H. (1972), “Fairly Good Plans”, *Journal of Economic Theory*, 4: 268–88.
Mueller, B. (2006), *Adaptation Funding and the World Bank Investment Framework Initiative*, unpubl. Background Report prepared for the Gleneagles Dialogue Government Working Groups, Mexico, 7–9 June.

NAESS, A. (1973), "The Shallow and the Deep, Long-Range Ecology Movement: A Summary", *Inquiry*, 16: 95–100.

NORDHAUS, W. D. (1994), *Managing the Global Commons: The Economics of Climate Change* (Cambridge, Mass.: MIT Press).

—— (2007), "A Review of the Stern Review on the Economics of Climate Change", *Journal of Economic Literature*, 45: 686–702.

NORTON, B. G. (1992), "Epistemology and Environmental Values", *Monist*, 75: 208–26.

NUNES, P. A. L. D., and VAN DEN BERGH, J. C. J. M. (2001), "Economic Valuation of Biodiversity: Sense or Nonsense?", *Ecological Economics*, 39: 203–22.

OKUN, A. M. (1975), *Equality and Efficiency: The Big Tradeoff* (Washington, DC: Brookings Institution).

RAWLS, J. (1971), *A Theory of Justice* (Oxford: Oxford University Press).

REES, M. J. (2003), *Our Final Century: A Scientist's Warning* (London: Heinemann).

ROLSTON, H., III (1975), "Is There an Ecological Ethic?", *Ethics*, 85: 93–109.

ROUTLEY, R. (1973), "Is There a Need for a New, an Environmental, Ethic?", *Proceedings of the 15th World Congress of Philosophy*, 1: 205–10.

SCHELLNHUBER, H. J., CRAMER, W., NAKICENOVIC, N., WIGLEY, T., and YOHE, G. W. (2006), *Avoiding Dangerous Climate Change* (Cambridge: Cambridge University Press).

SELDEN, L. R. (1979), "An OCE Analysis of the Effect of Uncertainty on Saving under Risk Preference Independence", *Review of Economic Studies*, 46: 73–82.

SEN, A. K. (1972), "Control Areas and Accounting Prices: An Approach to Economic Evaluation", *Economic Journal*, 82: 486–501.

—— (1992), *Inequality Re-Examined* (Cambridge, Mass.: Harvard University Press).

—— (1999), *Development as Freedom* (New York: Knopf).

—— and B. WILLIAMS (eds), (1982), *Utilitarianism and Beyond* (Cambridge: Cambridge University Press).

SHUE, H. (1999), "Bequeathing Hazards: Security Rights and Property Rights of Future Humans", in M. Dore and T. Mount (eds), *Global Environmental Economics: Equity and the Limits to Markets* (Oxford: Blackwell).

SINGER, P. (1977), *Animal Liberation: A New Ethics for our Treatment of Animals* (New York: Avon).

SOLOW, R. M. (1974), "Intergenerational Equity and Exhaustible Resources", *Review of Economic Studies Symposium*: 29–46.

STERN, N. (2007), *The Economics of Climate Change: The Stern Review* (Cambridge: Cambridge University Press).

TOL, R. S. J. (1997), "On the Optimal Control of Carbon Dioxide Emissions: An Application of FUND", *Environmental Modelling and Assessment*, 2: 151–63.

WEITZMAN, M. L. (2007), "A Review of the Stern Review on the Economics of Climate Change", *Journal of Economic Literature*, 45: 703–24.

WHITE, L., JR. (1967), "The Historical Roots of our Ecological Crisis", *Science*, 155: 1203–7.

WIGGINS, D. (1987), *Needs, Values, Truth* (Oxford: Blackwell).

—— (2006), *Ethics: Twelve Lectures on the Philosophy of Morality* (London: Penguin).

CHAPTER 20

HAS DEVELOPMENT AND EMPLOYMENT THROUGH LABOR-INTENSIVE INDUSTRIALIZATION BECOME HISTORY?

RIZWANUL ISLAM

I. Introduction

Ever since the publication of an influential article on economic development by W. A. Lewis (1954), economic development in countries with surplus labor has been described as a process of economic transformation towards more modern sectors

The views expressed in this chapter do not necessarily reflect those of the International Labor Office, to which the author is attached. Assistance provided by Irmine Iroko in compiling data for this paper is thankfully acknowledged. However, the usual disclaimer applies.

and the transfer of surplus labor from traditional sectors to modern sectors until such surplus is exhausted and real wages start rising. Indeed, development in a few countries of East Asia (viz. the Republic of Korea, Taiwan–China, Hong Kong and Singapore) during the 1970s and 1980s followed that general pattern. A number of countries of South-East Asia, especially Malaysia and Thailand, and to a lesser extent, Indonesia and the Philippines, followed a similar journey towards the so-called "Lewis turning point"—at least until they were hit by the economic crisis during 1997/8.

The broad contours of development in countries of East and South-East Asia (henceforth referred to as ESEA) mentioned above are quite well known by now. On the economic front, they all followed relatively more outward-looking and export-oriented strategies. High levels of private domestic investment—financed in large part through high levels of domestic savings, coupled with investment in human capital—acted as major engines of growth in all these countries (albeit in varying degrees). Labor markets were characterized by a high degree of flexibility in terms of responsiveness of wage rates to changes in demand for labor and the ability of firms to hire and fire without much intervention by either governments or trade unions. And it has often been argued that such flexibility in the labor market helped investment, economic growth and expansion of employment.[1] Rapid economic growth of an employment-intensive nature, in turn, led to impressive rates of poverty reduction.

The model of economic development described above became so influential that it was widely regarded as something to be emulated by other developing countries. However, as early as the eighties, questions were raised about the difficulty of replicating some of the initial conditions of success and the extent to which such a model could be emulated by other countries (Lee 1981). More than two decades have passed since then. Many countries in the developing world are characterized by surplus labor and are at various stages of development. They are often advised to pursue open (or export-oriented) economic and trade policies—the expectation being that such strategies will enable them to achieve development, absorb their surplus labor, and reduce poverty. It is therefore quite legitimate to raise the question as to the degree to which such countries with surplus labor have been able to achieve the kind of development achieved by the initial set of countries mentioned above. The question to ask, in more specific terms, is: to what extent have other developing countries with surplus labor been able to achieve economic growth with similar employment intensity and to move towards the Lewis turning point, as was the case in the countries that attained success in this regard? A related research question is to understand the factors that can explain the dissimilarities, if any,

[1] There is a large body of literature on the East Asian experience of economic development. While good accounts are provided by Lee (1981) and World Bank (1993), a more recent analysis can be found in Storm and Naastepad (2005).

between the success stories of the earlier decades and the more recent development experience.

The present chapter begins with a recapitulation of the East Asian experience from the perspective of the employment intensity of economic growth that was achieved (section II). It then undertakes, in section III, a comparative analysis of the degree of success achieved by various countries of East, South-East and South Asia in attaining the so-called Lewis turning point in terms of changes in the structure of their labor markets. Section IV looks at the recent experience of the developing Asian countries with respect to the employment intensity of their economic growth, and examines possible factors that could explain the differences in employment intensity of growth. The concluding section points out the difficulties faced by contemporary developing countries of Asia in achieving development and employment through labor-intensive industrialization.

II. Recounting the Experience of Labor-Intensive Industrialization and Development

Amongst the late industrializers, four countries of ESEA, namely Republic of Korea, Hong Kong, Singapore and the Taiwan province of China (henceforth referred to as Taiwan), are regarded as having succeeded in achieving development through export-oriented industrialization. However, in terms of possible emulation by other developing countries, the latter three are often considered to be non-representative because of their small size and other special characteristics. Therefore, this chapter will use the experience of Korea as a success story and see how far the other developing countries of Asia have been able to proceed along the path followed by that country.

On growth in the region as a whole, one should notice several features. First, Korea and some other countries of the ESEA region achieved high and sustained economic growth over several decades (Korea starting in the 1960s and some others from the 1970s). The rate of economic growth (measured by annual compound rate of GDP growth) has been consistently higher in the ESEA countries compared to the South Asian countries (see Table 20.1).

The second aspect, and an important one, of the growth experience (which also emerges from the data presented in Table 20.1) has been the high growth rate of manufacturing industries, especially in Korea and also in some other ESEA countries. Indeed, the elasticity of growth of manufacturing with respect to overall

Table 20.1. Growth rate of overall GDP and manufacturing output, South and East Asia (annual compound rate of growth, %)

Country	1960–70			1970–80			1980–90			1990–6			2000–5		
	GDP	Man	EMG	GDP	Man	EMG	GDP	Man	EMG	GDP	Man	EMG	GDP	Man	EMG
Bangladesh	n.a.	n.a.	n.a.	2.3	5.1	2.2	4.3	3.0	0.7	4.3	7.3	1.7	5.4	6.7	1.2
Cambodia	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	6.5	7.8	1.2	8.9	14.1	1.6
China	n.a.	n.a.	n.a.	5.5	10.8	1.96	10.2	11.1	1.1	12.3	17.2	1.4	9.6	11.1	1.2
India	3.6	4.8	1.3	3.4	4.6	1.4	5.8	7.4	1.3	5.8	7.5	1.3	7.0	6.9	0.99
Indonesia	3.5	3.3	0.9	7.2	14.0	1.9	6.1	12.8	2.1	7.7	11.1	1.4	4.7	5.2	1.1
Republic of Korea	8.5	17.2	2.02	10.1	17.7	1.8	8.9	12.1	1.4	7.3	7.9	1.1	4.6	7.0	1.5
Malaysia	6.9	n.a.	n.a.	7.9	11.7	1.5	5.3	9.3	1.8	8.7	13.2	1.5	4.8	5.2	1.1
Nepal	n.a.	n.a.	n.a.	2	n.a.	n.a.	4.6	9.3	2.0	5.1	12.0	2.4	2.8	−0.6	n.a.
Pakistan	6.7	9.4	1.4	4.9	5.4	1.1	6.3	7.7	1.2	4.6	5.5	1.2	4.8	9.1	1.9
Philippines	5.1	6.7	1.3	6.0	6.1	1.0	1.0	0.2	0.2	2.9	2.6	0.9	4.7	4.3	0.9
Sri Lanka	4.6	6.3	1.4	4.1	1.9	0.5	4.0	6.3	1.6	4.8	8.8	1.8	4.2	2.9	0.7
Thailand	8.2	11.0	1.3	7.1	10.5	1.5	7.6	9.5	1.3	8.3	10.7	1.3	5.4	7.2	1.3
Vietnam	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	4.6	n.a.	n.a.	8.5	n.a.	n.a.	7.5	11.5	1.5

Abbreviations: EMG: Elasticity of manufacturing growth with respect to GDP growth; Man: Manufacturing output; n.a. = not available.

Sources: World Bank (*World Development Indicators* 1998, 2004, 2007; *World Development Report* 1978, 1995).

Table 20.2. Growth of total exports and manufactured exports, South and East Asia (annual compound rate of growth, %)

Country	1960–70		1970–80		1980–90		1990–6		2000–5	
	Total exports	Manufactured exports[a]	Total exports	Manufactured exports[a]	Total exports	Manufactured exports[a]	Total exports[b]	Manufactured exports[a]	Total exports[c]	Manufactured exports[ad]
Bangladesh	6.6	n.a.	−2.4	n.a.	7.6	8.4	12.7	n.a.	8.6	7.2
Cambodia	−3.3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	6.7	n.a.
China	23.7	n.a.	8.7	n.a.	11.0	18.0	14.3	18.7	26.7	26.2
India	3.1	n.a.	5.9	n.a.	6.5	10.0	7	10.5	19.2	14.8
Indonesia	3.5	n.a.	6.5	n.a.	2.8	35.2	21.3	19.0	7.1	2.7
Republic of Korea	35.2	n.a.	22.7	n.a.	12.8	14.5	7.4	11.1	12.9	10.5
Malaysia	6.1	n.a.	3.3	n.a.	10.3	18.0	17.8	28.9	8.8[e]	6.1
Nepal	n.a.	n.a.	n.a.	n.a.	4.5	17.1	22.1	19.8	1.1	0.3
Pakistan	8.2	n.a.	3.1	n.a.	9.0	11.8	8.8	12.1	12.7	10.9
Philippines	2.2	n.a.	7.2	n.a.	2.5	9.7	10.2	21.2	1.7	0.1
Sri Lanka	4.6	n.a.	−1.4	n.a.	6.8	18.7	17.0	21.0	4.1	2.9
Thailand	5.2	n.a.	8.9	n.a.	13.2	23.2	21.6	18.5	11.1	10.1
Vietnam	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	17.7	n.a.

n.a. = not available.

[a] Manufactured exports rates are calculated with the values of exports, to which the formula $(y_t/y_o)^{1/t} - 1$ is applied.

[b] Data are taken from WDI 1997 and are for 1990–5.

[c] Data are taken from UNCTAD (2006).

[d] Calculations are made with data from WDI 2002 and 2007.

[e] Internal trade between the states of Malaysia included.

Sources: UNCTAD (2006); World Bank (*World Development Report* (WDR) 1979, 1992, 1995, 1997, 1998/9, 1999/2000; *World Development Indicators* (WDI) 1997, 2002, 2007).

GDP growth was over 2 in Korea in the 1960s and dropped only to just below 2 (1.9, to be precise) during 1970–80. It dropped below 1.5 only during the 1980s. In Malaysia also, the elasticity of manufacturing output growth with respect to overall GDP growth has been between 1.5 and 2 for almost three decades (during 1970–96). Indonesia and Thailand also had a similar experience. On the other hand, this figure has been in the range of 1.3 to 1.4 in India, and lower in Pakistan. In Bangladesh and Sri Lanka, there has been a considerable degree of fluctuation in the elasticity of manufacturing growth with respect to GDP growth. On the whole, it is quite clear that not only has overall economic growth been higher in Korea and a few other countries of ESEA (the Philippines being an exception) than in South Asia, the manufacturing sector has been the major (and a more important) driver of growth in the former. This has enabled Korea (and other countries of ESEA in varying degrees) to achieve structural change in their economies of a kind that is in line with the Lewis model. Of course, not all countries of ESEA succeeded in achieving such a transformation, especially of their labor markets—an issue to which we shall turn in the next section.

The third important aspect of economic growth in the countries of ESEA compared to South Asia has been the superior export performance of the former and the greater openness of their economies (Lee 1981; World Bank 1993). Indeed, the share of exports in total GDP has been consistently much higher in ESEA compared to South Asia. Even after the economic reforms introduced in the Indian subcontinent in the 1980s and 1990s and the gradual opening up of these economies, the share of exports in GDP in Bangladesh, India and Pakistan has remained below 25%, whereas it has ranged between 40 and 130% in ESEA (Islam 2003). More important from the point of view of industrialization, the growth rate of manufacturing exports has been much higher in ESEA compared to South Asia (Lee 1981; Table 20.2 above).

From the point of view of labor-intensive industrialization, it is important to look at the composition of industrial exports. In this regard, the experience of Korea in particular is noteworthy. During the 1960s (i.e. the early stages of export-led development) the share of "light industries" in manufactured exports remained high—ranging from 83% in 1964 to 89% in 1968. Only in the 1970s did this share begin to fall, declining to 60% in 1978 (Park 1981).[2] Between 1960 and 1970, direct employment in exports in Korea increased more than 18-fold, or at an average annual rate of 34%. As a result, the share of export industries in total employment expanded from 5 to 22.5% during the same period (Hsia 1981). In fact, the degree of capital intensity (measured by capital–labor ratio) of Korean manufacturing declined during the 1960s until the early 1970s (Little 1981).[3] The capital intensity of

[2] There is a body of literature documenting the labor-intensive character of Korea's export-led industrialization during the 1960s. For references, see Park (1981).

[3] Hong Kong and Taiwan also provides good examples of exports of labor-intensive manufactures. In Taiwan, the manufactured exports were more labor-intensive than manufacturing in general. During the early phase of its industrialization, Taiwan also utilized a variety of means (e.g. the use of

Korean manufacturing and of exports in particular started rising only in the early 1970s.

The labor-intensive character of Korea's early stages of industrialization is also illustrated by the high employment intensity of growth in the manufacturing sector. While estimates for the 1960s are difficult to come by, according to one estimate (Khan 2006), the elasticity of employment with respect to output growth in manufacturing was 0.69 during the 1970s, declining to 0.49 during the 1980s. It is thus clear that even after almost a decade of industrialization with high rates of output growth, the manufacturing sector in Korea remained highly employment-friendly. Moreover, there was a decline in the elasticity of employment only after surplus labor was exhausted. It should be added here that with an employment elasticity of 0.69, there was still considerable scope for labor productivity to improve; and that is what happened in reality. The high rate of growth-induced expansion in employment was accompanied by high growth in productivity as well as in the earnings of workers (Storm and Naastepad 2005; Khan 2006).

III. The Journey towards the Lewis Turning Point: The Varying Degrees of Success

III.1 Indicators of Success in Attaining the Lewis Turning Point

A critical point in the development of an economy with surplus labor is where the traditional sector (which is the reservoir of such labor) starts facing a shortage of labor, the modern sector is no longer able to hire workers without raising real wages, and the margin between the wage rates prevailing in the two segments of the labor market starts increasing. In analyzing the degree of success attained by an economy in absorbing its surplus labor through industrialization and development, it is important to examine whether it has reached such a point (which, as noted earlier, is sometimes referred to in the development literature as the "Lewis turning point"). But another important question is the criteria to be used in judging whether an economy has indeed reached that point.

In the original version of the Lewis model, the supply curve of labor remains horizontal at a given level of real wages until the surplus labor available in the traditional sector gets exhausted. Up to that point, the demand curve for labor

labor-intensive technology and even second-hand machines and multiple-shift operation) to make their production labor-intensive. Korea followed similar practices in the 1960s. See, for example, Little (1981) and Ranis (1973) for descriptions of the experiences of Taiwan and Korea respectively.

can continue to shift towards the right; thus the economy can employ more labor at current real wages and create a “surplus” for further accumulation until the economy reaches a point where real wages start rising. Thus the point at which real wages begin to rise could be regarded as an indicator of the turning point. But in reality it may be difficult to apply this criterion because real wages may start rising even before surplus labor is fully exhausted. This can be due to a variety of factors, e.g. the willingness of employers to share the gains of a productivity increase in order to maintain the level of productivity and a loyal labor force, or interventions by the government and trade unions. Indeed, in many developing countries with surplus labor—including the Republic of Korea when it still had surplus labor—real wages have demonstrated upward trends (albeit slowly at times), especially when labor productivity increased.

It is therefore necessary to look for alternative indicators of labor market tightening and attainment of the Lewis turning point. Candidates in this regard include: (1) the rates of open unemployment and underemployment, (2) the proportion and—more important—the absolute numbers of the labor force employed in the traditional sector (e.g. agriculture), and (3) the proportion of the labor force engaged in the so-called informal economy.

III.2 How Have the Developing Asian Economies Performed?

Korea, Hong Kong, Singapore and Taiwan were able to complete the Lewis transformation of their labor market quite successfully. In Korea, the unemployment rate declined steadily throughout the 1960s and real wages started rising sharply in the 1970s,[4] and the percentage of employment in manufacturing increased threefold between 1963 and 1978. While the proportion engaged in agriculture registered a steady decline throughout the sixties, the absolute number in that sector started falling after 1976.[5] On the whole, it can be said that Korea reached the Lewis turning point by about the mid-1970s.

Which other countries of Asia have achieved a degree of success similar to that of Korea? If trends in unemployment are considered (Table 20.3), it would appear that the labor markets in countries like Malaysia and Thailand became tighter in the 1990s (especially before the East Asian economic crisis in 1996/7). However, even on the basis of this relatively imperfect criterion, not many countries have been very successful. Indonesia, for example, did attain a low rate of unemployment before the economic crisis; but since then the open unemployment rate has risen to double figures. In the Philippines, unemployment in the 1990s and the first half of the present decade has been higher than in the eighties. In South Asian countries

[4] Real wages in Korean manufacturing were already rising in the 1960s, but the increase remained well below the increase in labor productivity during that period.

[5] The information in this paragraph is from Park (1981).

Table 20.3. Unemployment in selected Asian countries

Country and year	Unemployment rate (% of labour force)	Country and year	Unemployment rate (% of labour force)
Bangladesh		Nepal	
1984	1.8	1996	4.5
1991	1.9	2001	8.8
2003	4.3	Pakistan	
China		1980	3.7
1980	4.9	1990	2.6
1990	2.5	2000	7.2
2000	3.1	2005	7.7
2005	4.2	Philippines	
India		1980	4.8
1994	3.7	1990	8.1
2000	4.3	1996	7.4
2004	5.0	2000	10.1
Indonesia		2005	7.4
1982	3.0	Sri Lanka	
1992	2.8	1985	13.3
2000	6.1	1990	14.4
2005	10.3	2000	7.4
Republic of Korea		2005	7.6
1980	5.2	Thailand	
1990	2.5	1980	0.9
1996	2.0	1990	2.2
2000	4.4	1996	1.1
2005	3.7	2000	2.4
Malaysia		2005	1.3
1984	5.8	Vietnam	
1990	4.7	1996	1.9
1996	2.5	2000	2.3
2000	3.0	2004	2.1
2004	3.5		

Source: ILO (2006).

like Bangladesh, India and Pakistan, unemployment rates have been generally low (perhaps because of the way unemployment is measured), but have shown a tendency to rise in recent years. Underemployment has also remained high in South Asian countries as well as in Indonesia and the Philippines. In fact, in the latter two nations, the rates have gone up since the Asian economic crisis (Islam 2003).

Data on the sectoral composition of employment in South and East Asia (Table 20.4) show declines in the share of agriculture and increases in the share of manufacturing and services. However, there are considerable variations in the nature and pace of structural change in the composition of employment. First, except in Malaysia and Thailand, the rate of increase in the share of manufacturing

Table 20.4. Sectoral composition of GDP and employment, South and East Asia (%)

Country	1970			1980			1990			1995			2000			2005		
	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services	Agriculture	Industry	Services
Bangladesh																		
GDP	55[a]	9[a]	37[a]	34	24	42	38	15	46	31	18	52	25	24	51	21	28	52
Employment	n.a.	n.a.	n.a.	68.7[b]	7.9[b]	23.4[b]	66.4	13	16.2	n.a.	n.a.	n.a.	62.1	10.3	23.5	n.a.	n.a.	n.a.
Cambodia																		
GDP	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	51	14	34	37	20	42	33	29	38
Employment	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	73.7	8.4	17.7	60.3[c]	12.5[c]	27.2[c]
China																		
GDP	n.a.	n.a.	n.a.	30	49	21	27	42	21	21	48	31	16	51	33	13	46	41
Employment	n.a.	n.a.	n.a.	68.7	18.2	11.7	53.4	19	9.9	48.5	21	12.2	46.3	17.3	12.7	46.9[c]	22.5[c]	30.6[c]
India																		
GDP	45[a]	22[a]	33[a]	38	24	39	31	29	40	29	29	41	25	27	48	19	28	54
Employment	n.a.	n.a.	n.a.	71.0[b]	13.0[b]	16.0[b]	69.1	13.6	17.3	66.7	12.9	20.3	n.a.	n.a.	n.a.	54.0[c]	20.0[c]	26.0[c]
Indonesia																		
GDP	44.94	18.69	36.37	23.97	41.72	34.31	20.42	37.64	41.94	17.14	41.8	41.06	17	47	36	14	41	45
Employment	65.9	7.9[b]	26.2[d]	55.9	13.2/ 9.3[b]	30.2/ 34.5[d]	55.9	13.7	30.2	46.1	18.7	35	45.1	17.5	37.3	44	18	38
Republic of Korea																		
GDP	27.08	29.48	43.45	14.84	39.38	45.27	8.51	43.11	48.39	6.19	43.2	50.61	5	43	53	4	41	55
Employment	55[e]	15[e]	37[e]	34	29	37	17.9	35.4	46.7	12.4	33.3	54.3	10.6	28.1	61.2	7.7[f]	18.1[f]	74.2[f]
Malaysia																		
GDP	29.44	27.39	43.17	22.61	41.04	36.35	15.22	42.2	42.59	12.95	41.4	45.65	11	45	44	9	50	40
Employment	50.5	11.4[b]	38.1[d]	37.2	24.1/ 15.5[b]	38.7/ 47.3[d]	26	27.5	46.5	20	32.3	47.7	18.4	32.2	49.5	14.8[g]	30.1[g]	52.5[g]
Nepal																		
GDP	67[a]	12[a]	21[a]	62	12	26	60	14	26	42	22	36	40	22	37	40	21	38
Employment	n.a.	n.a.	n.a.	91.0[b]	4.0[b]	5.0[b]	83.3	2.3	13.7	78.5	5.5	21	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Pakistan																		
GDP	37[a]	22[a]	41[a]	30	25	46	26	25	49	26	24	50	26	23	51	22	25	53
Employment	n.a.	n.a.	n.a.	52.7[b]	14.6[b]	32.7[b]	51.1	19.8	28.9	46.8	18.5	34.6	48.4	18	33.5	n.a.	n.a.	n.a.
Philippines																		
GDP	29.52	31.65	38.83	25.12	38.79	36.1	21.9	34.47	43.62	21.63	32.06	46.31	16	31	53	14	33	53
Employment	53.7	11.9[b]	34.4[d]	51.8	15.4/ 11[b]	32.8/ 37.6[d]	45.2	15	39.7	44.1	15.6	40.3	37.4	16	46.5	37	14.9	48.1
Sri Lanka																		
GDP	28[a]	24[a]	48[a]	28	30	43	26	26	48	23	25	52	20	27	53	17	26	57
Employment	n.a.	n.a.	n.a.	44.5[b]	11.8[b]	43.7[b]	47.8	20.6	30	37.3	23.4	33.6	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Thailand																		
GDP	25.92	25.31	48.78	23.24	28.68	48.08	12.5	37.22	50.28	11.18	39.16	49.65	10	40	49	10	47	44
Employment	79.2	4[b]	16.8[d]	70.8	10.3/ 7.9[b]	18.9/ 21.3[d]	64	14	22	52.1	19.8	28.1	48.8	19	32.2	42.6[h]	20.2[h]	37.1[h]
Vietnam																		
GDP	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	28	30	42	24	37	39	22	40	38
Employment	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	65.3	12.4	22.3	57.9[c]	17.4[c]	24.7[c]

n.a. = not available.

[a] GDP and its components are at purchaser values.

[b] Figures are from ILO (1993), and represent manufacture only.

[c] Figures are for 2004.

[d] Figures from ILO (1993).

[e] Figures are for 1965.

[f] Figures are for 2006.

[g] Figures are for 2004. Other sectors are excluded and represent 2.6% of total.

[h] Other sectors are excluded and represent 0.1% of total.

Sources: ADB (*Key Indicators* 1991, 2002, 2007); World Bank (*World Development Report* 1995) for some data in 1970; ILO (1993, 2005, 2006).

Table 20.5. Agricultural employment in selected Asian countries, 1970–2005

Country	Employment in agriculture (millions)				Employment in agriculture (%)[a]			
	1970	1980	1990	2005	1970	1980	1990	2005
Bangladesh	11.90[b]	13.00	n.a.	22.90[c]	73.91	68.71	n.a.	51.69
Cambodia	n.a.	n.a.	4.014[d]	4.48[e]	n.a.	n.a.	81.38	70.00
China	n.a.	291.81	389.14	339.18	n.a.	68.89	60.10	44.73
India	n.a.	207.31[f]	239.41[g]	258.79[h]	n.a.	n.a.	69.10	54.00[h]
Indonesia	n.a.	28.83	42.38	41.81	n.a.	55.93	55.87	44.04
Rep. of Korea	4.92	5.15	3.24	1.82	50.45	38.43	17.90	7.94
Malaysia	1.75[i]	1.80[i]	1.74[i]	1.48[i]	52.55	42.53	25.99	14.72
Pakistan	10.13	12.64	15.43	18.60	57.03	52.64	51.14	43.04
Philippines	6.33	8.45	10.19	12.17	53.79	51.44	45.20	37.02
Sri Lanka	1.66[j]	1.88[kj]	2.36[j]	2.31[j]	50.41	45.55	46.78	30.67
Thailand[l]	12.55	15.94	19.73	15.45	78.34	70.78	63.95	42.56
Vietnam	n.a.	n.a.	21.20	24.30	n.a.	n.a.	72.11	56.91

n.a. = not available.
[a] Employment in agriculture divided by total employment.
[b] 1973.
[c] 2003.
[d] 1995.
[e] 2002.
[f] 1983.
[g] 1993/4.
[h] 2004/5.
[i] Forestry, fishing and livestock.
[j] Data exclude both Northern and Eastern Provinces.
[k] 1981.
[l] Prior to 1986, the labor force refers to population aged 13 years and over; thereafter, it is the population aged 11 years and over.

Sources: ADB (*Key Indicators*, 1976, 1980, 1991, 2005, 2006 (online); for India: (1) Sundaram and Tendulkar (2002); (2) 2004/5: author's calculations from Unni and Ravindran (2007).

has been rather slow. In fact, the decline in the share of agriculture has been associated in many instances with an increase in the share of services rather than manufacturing. In India, for example, the share of manufacturing in total employment increased from 11.24% in 1983 to only 12.20% in 2004/5. In Bangladesh, the share of manufacturing in total employment actually declined during the 1990s, from 17% of the total employed labor force in 1990/1 to 9.8% in 2002/3 (Bangladesh Bureau of Statistics, various years).

In terms of the absolute numbers engaged in agriculture (Table 20.5), Malaysia witnessed a decline during 1980–90. In Thailand, a sustained decline began after 1992 (Krongkaew *et al.* 2006).[6] In Indonesia, the absolute numbers in agriculture

[6] According to labor force survey data presented in Krongkaew *et al.* (2006), there was a decline in numbers during the late eighties as well, but there were fluctuations. Since 1993, the decline has been continuous except for 1997, during which there was a small increase (caused mainly by the economic crisis) that was reversed the following year.

Table 20.6. Trends in the proportion of employment in the informal sector, South and East Asia

Country	Time period	Trend
Bangladesh	Since the 1980s	Increase
Cambodia	After 2000	Increase
India	1970s, 1980s, 1990s	Increase
Indonesia	1970s, 1980s, 1990s	Slight decline followed by increase
Malaysia	Since the 1980s	Decline
Nepal	1970s, 1980s and 1990s	Increase
Pakistan	Since the 1980s	Decline followed by increase
Philippines	Since the 1970s	Increase followed by no change since 2000
Sri Lanka	Since the 1990s	No major change
Thailand	Since the 1970s	Decline
Vietnam	Since the 1990s	Decline

Sources: Data available in Amin (2002: Table 9); ILO (2006).

began to decline in the early 1990s, but the trend was reversed in 1998—the year of severe economic crisis (Islam *et al.* 2001). Only in recent years has there been a small decline in the total number employed in the sector, and the share of that sector has not yet started declining at the same rate as before the crisis. On the other hand, Bangladesh, Cambodia, India, Pakistan, the Philippines and Vietnam have continued to see increases in the total number in agriculture, although the share of the sector has been declining.

Turning to the proportion of employment in the informal sector as an indicator of the tightening of labor markets, it is very difficult to obtain a clear picture because of the lack of a uniform definition and a paucity of data. Attempts have been made to quantify the magnitude of the sector (or that part of the economy which can be ascribed as informal) by using surrogates, e.g. the proportion of the self-employed in non-agricultural employment, or the sum of the self-employed, day laborers and unpaid family workers. An ILO report (ILO 2002) shows that between 1980 and 2000, self-employment as a share of non-agricultural employment declined in East Asia (from 23 to 18%), remained unchanged in South-East Asia (around 33%) and increased in South Asia (from 40 to 50%). Going beyond these sub-regional figures, it is possible to provide a picture of the broad direction of change in selected countries by using data compiled and presented by Amin (2002) and more recent data compiled by the ILO (2006). As the measure used is a surrogate and the data available refer to different years for different countries, rather than trying to prepare a time series, I attempt here to put together a picture of broad trends in various countries (see Table 20.6).

The picture presented in Table 20.6 is in line with the broad conclusions that can be drawn on the basis of other indicators of labor market tightening as described

earlier. The proportion of the informal sector in total employment has registered a marked decline only in Malaysia and Thailand, and in Vietnam in more recent periods. Going into details, we may notice that in the case of both Thailand and Vietnam the proportion of self-employment has actually increased over time. But in both cases, the proportion of unpaid family workers has gone down considerably—producing a decline in the proportion accounted for by these two categories. This is clearly an indicator of increases in the availability of income-yielding employment opportunities in the labor markets, drawing on the reserve of surplus labor in the form of family workers. The same trend may be observed in the case of Korea in the 1980s and Malaysia since the 1990s.[7]

The broad conclusions regarding the journey of the developing Asian countries towards the so-called Lewis turning point can be summarized as follows. Apart from the front runners, Malaysia and Thailand appear to be the only countries that have attained a degree of transformation in their labor markets indicative of the exhaustion of surplus labor. A country which came close to that point is Indonesia; but its journey has been stalled (and perhaps reversed) since the economic crisis of 1997/8. The other countries, e.g. Bangladesh, India, Nepal, Pakistan, and the Philippines, are still quite distant from that point.

IV. Persistence of Surplus Labor, Employment Intensity of Economic Growth and the Recent Development Experience in Asia

IV.1 Levels and Trends in the Employment Intensity of Growth

The failure of a majority of developing Asian countries to absorb their surplus labor through expanding employment in modern sectors like manufacturing gives rise to the important question of how this can be explained. For an answer, one needs to look at the overall economic growth rates and the growth of the manufacturing sector as well as the degree of employment intensity of the growth that has been achieved.

Growth rates themselves could offer an important explanation for the slow progress made by a number of countries (especially those in South Asia) during the

[7] These observations are made on the basis of data available in ILO (2006).

sixties, seventies and part of the eighties. Data presented in Table 20.1 demonstrate major differences in both overall GDP growth and manufacturing output growth between ESEA and South Asia during those periods. Indeed, the growth strategies pursued by the two sets of countries were quite different. While the former had moved towards more outward-looking and export-led development strategies by the late seventies and early eighties, the countries of South Asia, by and large, continued to follow a more inward-looking strategy till the eighties.[8]

However, during the eighties and nineties, the countries of South Asia also started moving towards greater openness and more outward-looking economic regimes. Although the time lag and magnitude of the impact of such changes on economic growth varied from country to country, all have been able to achieve higher rates of economic growth since the nineties. Particular mention may be made of India, where the average annual growth of GDP (real GDP at factor cost) during 1981 to 2005 was 5.8% per annum compared to 3.5% during the first three decades after independence (EIU 2006). As Table 20.1 shows, the growth rate has been higher in recent years. Similarly, GDP growth rates in Bangladesh since the second half of the nineties have been higher than in the 1980s (Islam 2007). But higher growth in these countries has not resulted in real structural transformation in the labor market and a rapid absorption of surplus labor in modern manufacturing. Likewise, one needs to return to the case of Indonesia, which had been pursuing an outward-looking and export-oriented development strategy since the early eighties and where the journey towards the Lewis turning point was proceeding, albeit slowly, till the economic crisis of 1997/8. The labor market, which was severely jolted by the crisis of 1998, does not appear to have recovered fully even though economic growth has resumed.[9] And the turning point has remained elusive.

The slow rate of labor absorption in modern manufacturing in some countries of Asia, despite high rates of GDP and manufacturing output growth, brings us to the issue of the pattern and the employment intensity of growth. As mentioned earlier, economic growth in several Asian countries has not resulted in a transformation of the employment structure towards modern manufacturing. That, in turn, seems to have been due in part to a low elasticity of manufacturing output growth with respect to GDP growth (see Table 20.1 and the discussion in section II), and in part to the low employment intensity of growth.

A basic measure of the employment intensity of growth can be provided by the elasticity of employment with respect to output growth. This measure, of course, is based on the assumption that the quantity of employment depends on the amount of output produced. According to this measure, employment elasticity reflects the

[8] Sri Lanka began opening up earlier, in the late seventies.

[9] As is evidenced from the fact (and the discussion in section III) that the unemployment rate has climbed to double figures in recent years.

inverse of labor productivity.[10] While an elasticity higher than 1 implies a decline in productivity, an elasticity lower than 1 means that employment expansion is taking place alongside an increase in productivity. A rise in productivity leads to a reduction in employment elasticity. Thus, in the interest of combining employment expansion with rising productivity, it is important to have an employment elasticity of less than 1, but it need not be very low.[11]

As long as employment elasticity is not too low in a country with surplus labor and does not decline further from a low level, a country could have employment-intensive growth even when employment elasticity is less than 1 and is declining. The characteristics of a country experiencing employment-intensive growth along with improvements in labor productivity include high growth of output in modern sectors (e.g. manufacturing, construction, modern services), reasonably high initial levels of employment elasticity, and a gradual decline in employment elasticity, as surplus labor is employed in higher-productivity economic activities and is gradually exhausted. This is the pattern of development experienced by Korea in the 1960s and 1970s and the more successful developing Asian countries like Malaysia and Thailand in the 1980s and 1990s. Has this been happening in other developing countries of Asia? Estimates of employment elasticity (presented in Table 20.7) may be relevant to this question.

Although Table 20.7 presents estimates of employment elasticity for economies as a whole as well as for the manufacturing sector, it may be more appropriate to look at the latter because in developing countries, total employment is more likely to reflect the supply side of the labor force rather than demand for labor associated with output growth, and it is in sectors like manufacturing that employment can be expected to be a better reflector of the demand for labor. The figures presented in Table 20.7 enable us to draw important conclusions, albeit tentatively, regarding the employment intensity of growth in the manufacturing sector of the countries concerned.

First, there appears to have been a general decline in the employment intensity of growth in the manufacturing sector in the developing economies of Asia during the 1990s compared to the 1980s. While the declines in countries like Malaysia and Thailand could be taken as reflecting the tightening of their labor markets, the same cannot be said about other countries, which, as indicated earlier in this chapter, are still characterized by surplus labor. Second, during the 1980s, the level of employment intensity in manufacturing in some countries (e.g. Cambodia, China and India) was rather low in relation to the factor endowment. Indeed, those levels

[10] For a more rigorous demonstration of the relationship between employment elasticity and labor productivity, see Kapsos (2005).

[11] For a more detailed discussion of the conceptual difficulties associated with the concept of employment elasticity, and the upper limits of employment elasticity in an economy with surplus labor, see Islam (2006) and Auer and Islam (2006).

Table 20.7. Output elasticity of employment (OEE) in selected Asian countries

Country	OEE (economy-wide)		OEE (manufacturing)	
	1980s	1990s	1980s	1990s
Bangladesh[a]	0.55[b]	0.50[b]	0.76[c]	0.72[c]
Cambodia	n.a.	0.48	n.a.	0.56
China	0.33[b]	0.13[b]	0.50	0.25[d]
India	0.40	0.15	0.37	0.29
Indonesia	0.44[b]	0.38[b]	0.79[e]	0.61[f]
Malaysia	0.55	0.48	0.67[g]	0.71[g]
			0.55[h]	0.45[h]
Sri Lanka	0.51	0.46	0.64	0.47
Thailand	0.56	0.10	0.55	0.53

n.a. = not available.
[a] Islam (2006).
[b] Asian Development Bank (2005).
[c] Figures are based on data at three-digit level. Figures based on data at the four-digit level show a sharper decline, from 0.74 to 0.60. See Islam (2006: ch. 5).
[d] Figure is for 2002.
[e] Figure is for 1980–4.
[f] Figure is for 1990–4.
[g] UNDP and ILO (2007).
[h] Elasticity with respect to value added.

Sources: Unless specified otherwise, figures have been taken from recent ILO–UNDP country case studies, a synthesis of which can be found in UNDP and ILO (2007).

appear quite conspicuous if compared with the level in Korea (and also Malaysia) during the early stages of industrialization, when the economy still had surplus labor. Third, the degree of employment intensity seems to have declined further from such low levels.[12]

Low and declining employment elasticity in manufacturing in some countries of Asia implies that growth in those countries has been driven primarily by labor productivity. In that respect, countries like China and India (and to some extent Bangladesh, Cambodia and Sri Lanka also) appear to be pursuing a very different strategy of development and industrialization compared to that of ESEA. In countries like Korea, Malaysia and Thailand there has been a combination of productivity and employment growth—at least in their early stages of development (Storm and Naastepad 2005). From this perspective, it is important to understand how those countries have avoided a trade-off between growth in employment and labor productivity and why other countries are pursuing a strategy of growth primarily driven by productivity.

[12] These conclusions derived from Table 20.7 are very much in line with the data presented in Khan (2006).

IV.2 Factors Influencing the Employment Intensity of Economic Growth

Employment elasticity can change due to either a change in the composition of a sector or a change in the technology employed in a particular activity. The elasticity effect can therefore be seen as the sum of the "sector composition effect" and the "technology effect".[13] Each of these effects can be influenced by different sets of factors. For example, the sector composition of output in a particular country can be influenced by the comparative advantage of a country *vis-à-vis* its resource endowment, the overall policy environment, relative factor prices, characteristics of the labor market institutions, and the like. The technology effect, on the other hand, may be influenced by the availability of alternatives, relative factor prices, competitive pressures, and a host of other factors.

As for the policy environment, the nature of a trade regime can influence the pattern of industrialization in terms of its sector composition. One prediction which follows from the standard theory of international trade is that greater trade openness should lead to a rise in the employment intensity of output growth.

A related factor that is important from the point of view of overall policy environment is economic reforms. Alongside trade liberalization, many developing countries in Asia have been undertaking economic reforms aimed at moving towards more market-oriented economies. Such reforms, by removing distortions in various markets, including the factor markets, should lead to a more efficient allocation of resources, including factors of production. That, in turn, should result in the more efficient use of labor, the abundant factor of production in countries with surplus labor.

Relative factor prices (and hence the employment intensity of production) can also be influenced by interventions in the factor markets. On the one hand, governments in some countries, in an effort to provide incentives for investors, introduce measures that in effect make capital cheaper than it would otherwise have been. On the other hand, interventions in the labor market (e.g. through legislation, including laws on minimum wages) may raise the price of labor above what standard theory would regard as a market-clearing wage. Either way, such interventions could create distortions in factor markets which, in turn, could influence the degree of employment intensity of growth.

Although there is not much empirical evidence to subject the above hypotheses to rigorous tests, the available literature can shed some light, at least tentatively, on this issue.[14]

First, to judge from the cross-country regressions carried out by Kapsos (2005), trade openness does not appear to be statistically significant in explaining the

[13] See Osmani (2006) for a methodology for decomposing these two effects.

[14] The literature in this field is nascent; Kapsos (2005) provides a brief overview of some of the studies available.

degree of variation in the overall employment intensity of economic growth. This is counter-intuitive, especially if one considers developing countries expected to have comparative advantage in more labor-intensive items: trade openness is supposed to spur faster growth of sectors producing them, thus leading to higher employment intensity. Given such a hypothesis, one possible explanation for the result obtained by Kapsos (2005) could be that his regressions include both developed and developing countries. But on the other hand, even in developing countries, trade liberalization may not lead to a net expansion of employment if the number of jobs destroyed in enterprises going out of business because of competition exceeds the number of jobs created in new industries.

A study covering ten countries of Asia (Osmani 2006) finds that in the 1980s and 1990s, the weights of more labor-intensive industries in the manufacturing sector declined. A recent study on India's organized manufacturing sector (Palit 2007) brings out some interesting facts. While some labor-intensive sectors like textiles (including garments), footwear and jewelery have been able to achieve higher growth since the introduction of economic reforms in 1991, no discernible shift has taken place in the structure of industries. In fact, the share of the top five labor-intensive sectors (food products, beverages and tobacco, leather products, wood products and furniture, and other textiles) declined between 1990/1 and 2003/4. Moreover, while growth in some labor-intensive sectors has been negative, a good percentage of the relatively more capital-intensive industries has achieved faster growth. Thus, greater trade openness has not altered the structure of manufacturing towards more labor-intensive sectors.

These findings are also counter-intuitive in that during the 1990s the countries in question were more open to external trade than during the earlier decade, and so, according to the predictions of standard trade theory, one would expect the weights of more labor-intensive industries to increase rather than decline. This implies that the conventional wisdom regarding the expected impact of trade openness on employment intensity of growth needs to be looked at carefully.

A study on Indonesia (Islam and Chowdhury 2006) also indicates the limited impact of exports of labor-intensive manufactures and of trade openness on employment intensity of growth. In that country, the share of "high-technology" sectors in total exports increased during 1993 to 2002, while the share of "low-technology" sectors declined over the same period. Moreover, some of the more labor-intensive sectors, e.g. garments, furniture and rubber products, showed negative employment elasticity during the period 2000–3, implying that even within industries, changes are taking place (either in the technology used, or in the rationalization of the workforce through other means) that are affecting their employment-generating capacity. Thus there seems to be a combination of "sectoral composition effect" and "technology effect". In Malaysia, on the other hand, technological change (i.e. the adoption of more capital-intensive technology) is the most important factor explaining the decline in observed employment elasticity during the 1990s. That,

in turn, must be due to the change in the relative prices of the important factors (i.e. capital and labor) as the labor market gradually tightened (Nair, *et al.* 2006).

As mentioned earlier, market-oriented economic reforms may create conditions for employment-intensive economic growth in countries with surplus labor. However, the opposite may happen in reality, at least in the short and medium term. An example of such a perverse effect is provided by China, a country that has been instituting economic reforms for nearly three decades. Reform of state-owned enterprises has made it difficult for them to retain an army of "surplus labor", which they could support in the earlier system. As they started shedding their excess labor without necessarily having to curtail their output, the net result has been a sharp decline in the elasticity of employment with respect to output.[15]

Economic reforms also do not necessarily lead to a removal of distortions in factor markets. A good example of this is provided by various measures in India with the express purpose of providing incentives for investment. These measures, which include subsidies on investment and interest rates,[16] almost certainly make capital cheaper than it otherwise would have been. In the absence of any such measures to encourage the use of labor, the relative price of the two critical factors appears to favor the use of capital, although the opposite should be the case.

As for the importance of labor-market-related variables in explaining variations in employment elasticity, cross-country regressions in Kapsos (2005) indicate that labor market regulations are not a statistically significant explanation of differences in elasticities.[17] This is interesting, because it goes against the widely held notion that employment protection legislation reduces the demand for labor. A study of India by Mitra and Bhanumurthy (2006) also fails to find a significant relationship between employment and wages. Likewise, the study on Malaysia mentioned above (Nair *et al.* 2006) finds that the relationship between employment and real labor cost is statistically insignificant.

The literature on factors influencing the employment intensity of growth is still in its infancy. But this brief overview of findings from available studies indicates that it is difficult to draw clear conclusions on possible factors. The conventional wisdom does not often stand up to the test of reality. More research is needed in this area.

[15] However, as Khan (2006) has pointed out, it is quite possible that the actual labor input (measured in terms of labor time used) may not have declined to the extent indicated by the reduced number of workers employed, and thus the "observed decline" in employment elasticity may in fact overestimate the true fall in the employment intensity of growth. Once all of the excess labor has been shed, the decline in observed employment intensity should come to a halt.

[16] See Palit (2007) for more details on the various measures.

[17] That study uses the World Bank's employment rigidity index (which includes rigidity of hours, difficulty of hiring and firing, and firing costs) as the independent variable to represent labor market rigidity.

V. Concluding Observations

Outside the four pioneers, viz. Hong Kong, Korea, Singapore and Taiwan, success in achieving development and the absorption of surplus labor through labor-intensive industrialization has remained limited to a small number of developing Asian countries. How does one explain this? One well-known answer is that the laggards had been pursuing inward-looking economic policies that constrained their prospects of growth. However, this argument may no longer be as valid as it was during the sixties, seventies and the early part of the eighties. Most of the developing countries of Asia have now switched over to more outward-looking strategies of development and have been undertaking market-oriented economic reforms. Some have also been able to narrow the differences in the rate of economic growth. In fact, China during the last three decades and India for more than a decade have achieved growth rates comparable to those of the successful countries. Others have also succeeded in getting out of the low-growth regime. And yet, success in achieving the kind of industrialization and labor absorption in the pioneer countries has been elusive.

While the less successful countries do not belong to a single category, one characteristic in common is the low (and in some cases falling) employment intensity of growth in manufacturing output. There are cases (e.g. Bangladesh and Cambodia) where success has been confined to the growth of a single export-oriented industry. A somewhat special case is Indonesia, whose experience during the 1980s and the first half of the 1990s was one of success, but which has not been able to get back on track since it was hit by the economic crisis in 1998.

In an increasingly globalizing economy, a key concern is competitiveness and productivity. The four pioneers and their followers were able to combine growth in productivity with high employment growth, and that is the major difference between them and the less successful nations. In order to construct a strategy for development which combines high rates of output and employment growth, the key is to understand the reasons for this difference. Trade openness and economic reforms leading to a liberalized economy do not automatically result in an ability to combine productivity and employment growth.

Difficulties in replicating some of the conditions of the success of the front runners in labor-intensive and export-oriented industrialization (e.g. prior industrialization, the ability of the state to administer complicated incentive schemes, and flexible labor markets) were noted by some commentators in the early 1980s. The possibility of late starters facing greater competition in the international market or in attracting foreign investment was also noted (e.g. Lee 1981). And if emulating the success of the front runners was already difficult in the eighties, it must be all the more difficult now. The global context for trade and industrialization has changed substantially in various respects. First, competition based on cheap

labor has become much more intense with the emergence of new late starters like Bangladesh, Cambodia, Laos and Vietnam. In addition, large countries like China and India not only have the advantage of cheap labor but also fulfil one of the basic preconditions for success, namely a prior degree of industrialization. Given this scenario, the quest for productivity growth requires a greater focus on technology (and hence, capital) than was the case with front runners like Korea and Taiwan. Instead of a declining capital–labor ratio (Korea in the sixties) and the use of lower-level technology, including second-hand machines (e.g. in Taiwan), the recent experience has been one of deepening capital and continuous efforts to be at the frontier of technology. However, even countries with a stronger initial industrial base have not been able to achieve the same level of elasticity of manufacturing-output growth with respect to GDP growth as was achieved by Korea or Malaysia. Moreover, the degree of employment intensity has also been lower (and has been falling).

The second respect in which the global context has changed is the functioning of labor markets. What has been referred to as "superior labor market performance" in East and South-East Asia (World Bank 1993: 242) basically involves labor repression, either in the form of wage repression (euphemistically referred to as "wage restraint") or repression of rights and absence of union power. In the current global context of greater union power and international pressure for compliance, both seem to have become more difficult. If labor market flexibility implies labor repression of either kind, it is neither desirable nor feasible.

The overall policy environment may tilt further in favor of excessive use of capital when measures that are introduced in the name of creating incentives for investment artificially lower the price of capital (in India, for example). In the absence of any such incentives for the use of labor, the relative factor price gets distorted, resulting in the excessive use of capital (in the form of investments in capital-intensive products or technologies). Given the changed international context mentioned above and the policy environment prevailing in the economies of late starters in labor-intensive industrialization, it is not surprising that such countries have been much less successful than the front runners.

References

Amin, A. T. M. N. (2002), *The Informal Sector in Asia from the Decent Work Perspective*, Working Paper on the Informal Economy 4, Employment Sector (Geneva: ILO).

Asian Development Bank (2005), *Labour Markets in Asia: Promoting Full, Productive and Decent Employment* (Manila: Asian Development Bank).

Auer, P., and Islam, R. (2006), "Economic Growth, Employment, Competitiveness and Labour Market Institutions", in World Economic Forum (ed.), *The Global Competitiveness Report 2006–2007* (New York: Palgrave Macmillan).

Bangladesh Bureau of Statistics (various years), *Labor Force Surveys* (Dhaka: Bangladesh Bureau of Statistics).

Economist Intelligence Unit (2006), *Country Profile India* (London: *The Economist*).

Hsia, R. (1981), "Technical Change, Trade Promotion and Export-Led Industrialization", in Lee (1981).

ILO (International Labour Organization) (1993), *Papers and Proceedings of the Fifth Meeting of the Asian Employment Planners* (New Delhi: ILO).

—— (International Labour Office) (2002), *Women and Men in the Informal Economy: A Statistical Picture* (Geneva: ILO).

—— (2006), *Key Indicators of Labour Markets*, 4th edn (Geneva: ILO).

—— (2005), *Labour and Social Trends in Asia and the Pacific, 2005* (Bangkok: ILO).

Islam, I., and Chowdhury, A. (2006), *Growth, Employment and Poverty Reduction: The Case Study of Indonesia* (Geneva: ILO; Colombo: UNDP Regional Centre).

Islam, R. (2003), *Labour Market Policies, Economic Growth and Poverty Reduction: Lessons and Non-Lessons from the Comparative Experience of East, South-East and South Asia*, Issues in Employment and Poverty Discussion Paper 8 (Geneva: ILO).

—— (2004), *The Nexus of Economic Growth, Employment and Poverty Reduction: An Empirical Analysis*, Issues in Employment and Poverty Discussion Paper 14 (Geneva: ILO).

—— (ed.) (2006), *Fighting Poverty: The Development–Employment Link* (Boulder, Colo. and London: Lynn Rienner).

—— (2007), "What Kind of Economic Growth is Bangladesh Attaining?", paper presented at an international seminar on "Development Prospects of Bangladesh: Emerging Challenges" organized by the Bangladesh Institute of Development Studies in Dhaka on 2–3 December 2007 to celebrate its golden jubilee.

—— *et al.* (2001), "The Economic Crisis: Labour Market Challenges and Policies in Indonesia", in G. Betcherman and R. Islam (eds), *East Asian Labour Markets and the Economic Crisis: Impacts, Responses and Lessons* (Washington, DC and Geneva: World Bank and International Labour Office).

Kapsos, S. (2005), *The Employment Intensity of Growth: Trends and Macroeconomic Determinants*, Employment Strategy Paper 12 (Geneva: ILO).

Khan, A. R. (2006), "Employment Policies for Poverty Reduction", in Islam (2006).

Krongkaew, M., Chamnivickorn, S., and Nitithanprapas, I. (2006), *Economic Growth, Employment and Poverty Reduction Linkages: The Case of Thailand*, Issues in Employment and Poverty Discussion Paper 20 (Geneva: ILO).

Lee, E. (ed.) (1981), *Export-Led Industrialization and Development* (Bangkok: ILO).

Lewis, W. A. (1954), "Economic Development with Unlimited Supplies of Labour", *Manchester School of Economic and Social Studies*, 22: 139–91.

Little, I. M. D. (1981), "The Experience and Causes of Rapid Labour-Intensive Development in Korea, Taiwan Province, Hong Kong and Singapore, and the Possibilities of Emulation", in Lee (1981).

Mitra, A., and Bhanumurthy, N. R. (2006), *Economic Growth, Employment and Poverty: A Study of Manufacturing, Construction and Tertiary Sectors in India* (Geneva: ILO; Colombo: UNDP Regional Centre).

Nair, S., Yee, L. K., and Ching, G. S. (2006), *Economic Growth, Employment and Poverty Reduction: The Case Study of Malaysia* (Geneva: ILO; Colombo: UNDP Regional Centre).

Osmani, S. R. (2006), "Employment Intensity of Asian Manufacturing: An Examination of Recent Trends", unpubl. paper prepared for the Asian Regional Bureau of the UNDP, New York.

Palit, A. (2007), "Policies Responsible for Excessive Use of Capital Relative to Labour in India's Manufacturing Industries", unpubl. paper, Employment Sector, ILO, Geneva.

Park, Y. C. (1981), "Export-Led Development: The Korean Experience 1960–78", in Lee (1981).

Ranis, G. (1973), "Industrial Sector Labour Absorption", *Economic Development and Cultural Change*, 21: 387–408.

Sen, A. (1975), *Employment, Technology and Development* (Oxford: Clarendon Press).

Shyam, S. (2005), "Labour Flexibility Debate in India: A Comprehensive Review and Some Suggestions", *Economic and Political Weekly*, 28 May–4 June: 2274–85.

Storm, S., and Naastepad, C. W. M. (2005), "Strategic Factors in Economic Development: East Asian Industrialization 1950–2003", *Development and Change*, 36: 1059–94.

Sundaram, K., and Tendulkar, S. (2002), *The Working Poor in India: Employment-Poverty Linkages and Employment Policy Options*, Issues in Employment and Poverty Discussion Paper 4 (Geneva: ILO).

UNCTAD (United Nations Conference on Trade and Development) (2006), *Handbook of Statistics, 2006* (Geneva: UNCTAD).

UNDP (United Nations Development Programme) and ILO (International Labour Office) (2007). *Asian Experience on Growth, Employment and Poverty: An Overview with Special Reference to the Findings of Some Recent Case Studies* (Colombo: UNDP Regional Centre; Geneva: ILO).

Unni, J., and Raveendran, G. (2007), "Growth of Employment (1993–94 to 2004–05): Illusion of Inclusiveness?", *Economic and Political Weekly*, 42: 3 (20 January): 196–9.

World Bank (1993), *The East Asian Miracle: Economic Growth and Public Policy* (New York: Oxford University Press).

CHAPTER 21

IMPOSED ENVIRONMENTAL STANDARDS AND INTERNATIONAL TRADE

ROBERT M. SOLOW

A passing skeptic once asked Paul Samuelson to state a proposition in economics that was true and not obvious. Samuelson's spur-of-the-moment candidate was the theory of comparative advantage. Excellent choice, as your waiter or waitress is now likely to say. That, with the accompanying theorem about gains from trade, explains why nearly all economists favor generally open trade. Hardly anyone else does, except the narrowly self-interested. It would be an interesting project in political economy to understand why there is in fact so little protection when there is so much latent protectionism.

A fairly recent strategy that has turned up in US debates over regional free trade areas is to use the threat of cancellation or withdrawal to impose extraneous policies on other participants. Those others are usually poor countries (Latin America, Central America). The idea is to insist that they adopt higher labor standards or

An earlier version of this chapter formed the basis for a Centennial Lecture in November 2000 at the Stern School of New York University.

environmental standards than they otherwise would—thus “leveling the playing field”—or else the US will void a potential free-trade agreement, and deny access to its markets to the miscreants. The question is: how should a right-thinking person, like you or me or Amartya Sen, react to such a proposal? By a right-thinking person, I mean someone who would like to see the inhabitants of poor countries improve their standard of living, who would like to see low-wage workers better treated, and who generally favors environmental protection and free trade. When I thought about this question like a perfectly conventional economist, I came upon a rather unexpected conclusion, surprising not so much for its content as for its definiteness. This outcome made me think it would be a suitable tribute to an old friend.

There is another reason to be interested in this issue. Economists, as I have said, are generally in favor of free trade or open trade, for reasons that have been discussed and refined for at least 200 years. In this instance, however, the arguments offered, even by the most able and forceful advocates of free trade, seem to me not quite able to clinch the case. Usually what is said is that free trade is so valuable to poor countries as well as to rich ones (which is true) that it should not be held hostage to complaints about “environmental dumping”. Trade is one thing; the environment is another. Let us continue to have the benefits of free trade while we try to talk poor countries (and some richer countries, too) into more environmentally friendly policies. “Talk” can of course mean “subsidize” in this context.

This argument seems to me to have two kinds of weaknesses. I think it is a valid argument and a pretty good argument, but I also think that in polemical terms, in rough-and-tumble terms, it has some weaknesses. It is analytically weak because it ignores the basic fact that environmental policy has broader economic consequences, and general economic policy may often have broader environmental consequences. The two are not so easily separated. We really need to have an idea about the costs and benefits to poor countries of any proposed pattern of policy.

More important, I think that the standard argument for free trade without any attempt to impose environmental restrictions is pragmatically weak because it carries no weight with people who say, “We want to fight for the environment with every bit of leverage we can exercise, and trade issues provide some leverage. Why should we give that up in exchange for a vague and non-enforceable promise?” I want to see if there is an argument that is just as sound as the standard free-trade argument and harder to brush aside.

First, you should understand the nature of the argument that I plan to make. One of my goals is to illustrate the usefulness of general economic reasoning. I start by doing what any modern student of economics would do: make a small, simplified, analytical model of the situation that we need to evaluate. Think of it as an abstract toy economy that you can interrogate easily. It would be much too boring to try to explain the economic model in detail, and it is unnecessary because there is nothing original in it; that is part of its virtue. But you have to know what

pieces of economic reasoning and economic machinery produce the results, and that I will try to explain in plain language.

You are to imagine a rich country and a poor country trading with one another. I will not worry about the possibility that the rich country has some monopoly power; it may, but it may have less than you might think in the multi-country trading world of reality. Anyhow, it would not matter much for the points I want to emphasize. So, there are these two countries, a rich country and a poor country. Each country produces some stuff, exports some of what it produces, keeps the rest, imports some of the other country's output, and consumes.

In addition, each country's productive activity brings environmental disturbance. For brevity, I will just say that production causes pollution. How much pollution results from production is a matter of regulation, of voluntary or, more significantly, legally required and costly pollution-abatement activity.

There is an important distinction between local and global pollution. If pollution in the poor country physically affects the rich country, or vice versa—the standard examples are greenhouse gases and acid rain—then that is a suitable subject for international negotiation and bargaining. No one doubts that. It is, however, not a simple matter in that case to decide what the obligations of rich and poor countries ought to be, as we can see from the stand-off in connection with the Kyoto Agreement over greenhouse gases. But that is not something I have to discuss, because the business I have is with local pollution. The question raised by protests against the WTO or against environmental dumping is whether a rich country, like the US or the European Union, is doing the right thing if it closes its markets to goods produced in poor countries with low environmental standards, even if the resulting pollution affects only the poor country itself, like the bad air quality over Mexico City. So, let us imagine for the sake of the argument that all pollution is local; it does not spill over into the other country.

Of course no one pollutes for the sheer pleasure of damaging the environment. Most pollution occurs because avoiding it and preserving the environment is costly. For the purpose of analysis, I will make two reasonable assumptions about the cost. First, the higher the environmental standards that are enforced by regulation or custom, the more of its income a country must spend on pollution abatement. It is generally the case that each successive improvement in environmental standards is more costly to achieve than the previous one, so that perfection would be very expensive indeed. The second assumption is that once environmental standards are fixed, the amount of pollution is proportional to production. So, other things equal, a big country pollutes more than a small one.

The point of all this is just to remind you that each country faces a *trade-off*: if it wishes to improve its local environmental quality, it will have to spend more of its annual production on abatement. It will therefore have less left over to consume or export, and therefore it will have less capacity to import. Its population can have a better environment only by reducing their consumption of domestic and imported

goods. It seems natural to suppose that each country chooses the combination of material consumption and environmental amenity that fits its own preferences, political power relations and traditions best. (Within rich and poor countries alike, some people and groups would prefer more pollution abatement, and some would prefer less. That is not really relevant here. No one seriously advocates imposing environmental standards on poor countries as a way of shifting their political balance.)

One very common generalization is that very poor countries have very low consumption (because they are poor) and very little pollution (because industrial production is very low). I am not at all sure that this is always true, but it is the standard assumption. Then as income rises, so does pollution, presumably because those countries like the added consumption brought by faster industrialization more than they dislike the accompanying environmental degradation. Once some fairly high income level is reached, further increases in income result in lower pollution because consumption needs are being better met and a larger fraction of extra income is spent on improving the environment.

All that matters for my purpose is that at any moment a national economy could choose to consume more goods by accepting a worse environment or could choose to improve its environment in exchange for a lower level of consumption. Complaints about environmental dumping arise because poor countries choose to live with more pollution than protestors in Seattle or Genoa think is proper or ethical.

Now I have only two more details to mention about this toy economy. Each country produces its characteristic bundle of goods according to its own technology, using labor and capital as factors of production. The labor comes from its own population, and I would simply suppose that there is no unemployment, which would play no role in this kind of story anyway. Capital, however, is internationally supplied. In other words, I am assuming that labor is not mobile across borders, but capital is. A more realistic, but essentially equivalent, formulation is to say that labor moves across borders in response to wage differences much more slowly and with greater difficulty than capital moves across borders in response to differences in profitability.

The practical meaning of this presumption is that wage differences can persist between the poor country and the rich country, but that invested capital must earn more or less the same (risk-adjusted) return in both places. That is an extreme formulation. But it comes close to being true if we are thinking about capital invested in the manufacturing-export sector of each economy.

The last detail is that I will assume that trade (or the current account—it doesn't matter in this simple story) is balanced between the rich country and the poor country. That is not an essential assumption. Almost anything will do, as long as it is concrete and definite. For example, one could allow for a specified flow of

foreign aid from the rich country to the poor one. But balanced trade is just a neat simplification.

I said that the advantage of dealing with a simplified-model economy is that you can interrogate it. So, what question do we want to ask?

What I have in mind is the following. Imagine that these abstract economies have reached some kind of equilibrium. They are employing all their labor. Capital has flowed between them until the rate of profit is just about the same in both countries. Each country has imposed environmental regulations according to its own preferences for consumption and environmental amenity. Each country produces, exports, imports and consumes. Trade between them is in balance. In this equilibrium we can calculate many things, including the wage earned in each country.

Now the rich country says to the poor country, in the nicest possible way, "You pollute too much. In the interests of humanity we are going to insist that you impose stricter environmental regulations on your own citizens. If you do not, we shall feel obliged to exclude your exports from our markets. This will impose a hardship on you. But it is for a greater good." Faced with this situation, the poor country has no option but to reduce its pollution and accept the resulting costs. And now comes the question, "What else happens?" There has been a change in the underlying conditions. A new equilibrium situation must emerge. What can we say about it?

I am going to begin by stating the main answer, and then I will try to explain the economic machinery that makes it come out that way. The details of the argument will seem shorter if you know in advance where we are going.

In principle, when that initial equilibrium is disturbed, everything that can change will change. All the parts of these two toy economies are connected with one another. When one part moves, they all move. To say that is uninteresting. Much more interesting is one central conclusion: the (real) wage in the poor country will fall. When a new equilibrium is established with the stricter environmental regulation in the poor country, the workers in the poor country will be poorer in goods than they were before. To put it bluntly, the environmental conscience of the rich country will be satisfied at the expense of the poor country's workers. I do not suppose that is what demonstrators in Seattle want, but that is the likely outcome if they get their way on this issue. It's an example of *the law of unintended consequences.*

Now, having stated the conclusion that I come to, I want to explain the economic logic of this conclusion. There are (at least) three easily understandable pathways that lead in the same direction. The first is the most obvious. The reduction of pollution is costly. If it were not costly, there would be no environmental "problem", except, perhaps, out of sheer ignorance. When the poor country is compelled or induced to raise its environmental standards, it necessarily becomes less productive of ordinary goods and services. The same amount of labor and capital will create

fewer marketable goods. That is why we need legislation to enforce environmental standards.

But real wages or, better still, product wages—that is, wages measured in produced goods, rather than in money—are directly related to productivity in any market economy. An hour of labor is worth less to an employer if its net product is less. Imposing environmental costs on a firm is from this point of view exactly like making the firm less productive. The demand for labor in the poor country will fall, and the inevitable result is that wages will fall as well. It is not quite inevitable. We could go outside the model and imagine that wages were fixed. But then employment will fall. For simplicity, I am assuming that the labor market is flexible enough in the poor country so that employment is preserved, but with lower wages. The basic result is the same either way. Air quality is improved, but at the expense of a lower material standard of living for the vast majority. The lower standard of living can take the form of lower wages or lower employment, or both.

There is another minor qualification. I have assumed that the cost of pollution abatement is just the diversion of some amount of the poor country's normal bundle of outputs. If pollution abatement were an especially labor-intensive activity, an enforced shift from normal production to environmental improvement could conceivably induce a net increase in the demand for labor. There is no obvious plausibility to this twist. And anyway, the path to environmental standards is not aimed at beefing up the Civilian Conservation Corps.

That is not the end of the story. There is another chain of cause and effect to follow. When local industry becomes, to all intents and purposes, less productive, the earnings of capital will be reduced exactly like the earnings of labor and in just about the same proportion. But the symmetry ends there. In general, capital is a mobile factor of production, and labor is not. Of course, that is an exaggeration. Financial capital can move quickly and easily, but it may take time and trouble to extricate real capital from an unprofitable use. On the other hand, workers and their families do sometimes emigrate, legally or illegally, from low-wage countries to high-wage countries.

Nevertheless we expect capital to be significantly more mobile than labor. Another way to see this is to ask yourself if you would be more surprised to see capital earning much less in one place than another or to see labor earning much less in one place than another. We are accustomed to long-persisting wage differentials from place to place. We are much less accustomed to seeing persistent profit differentials, especially in manufacturing and export industries. I modeled capital as perfectly mobile, so that it earns the same return everywhere. That is extreme but it points in the right direction.

So the new chain of causation is this. Tougher environmental regulation is a direct cause of lower profits in the poor country. Capital will then, in time, leave the poor country to seek higher profits elsewhere. This process can be expected to

continue until capital has become scarce enough in the poor country to command the worldwide rate of return there. At the end of the process, production in the poor country will be less capital-intensive than it was at the beginning. That means the (marginal) productivity of labor will be further reduced, and the real wage in the poor country will have to fall further. The general principle illustrated here is that the mobile factor of production escapes the cost of regulation; the readily immobile factor is trapped. (In addition, the flight of portfolio capital from the poor country will depreciate the exchange value of the poor country's currency, and the resulting higher domestic costs of imported goods is another force depressing the standard of living.)

I will not follow that causal chain further. But there is one more. There is a third pathway of influence that can be distinguished from the first two, although it is closely connected with them. It deserves separate analysis because it specifically involves trade. In the toy economy, the rich country imports and consumes the poor country's exports, and the poor country imports and consumes the rich country's exports. I have made it a condition of equilibrium that trade should balance. But as I said, I could have allowed for other possibilities.

The poor country must export; it needs to be able to buy imports. Stricter environmental regulation tends to push up the cost and therefore the price of its export goods. The poor country's exports risk being priced out of rich-country markets, and this risk is greater if they must compete with close substitutes produced in other countries. So there is pressure for the poor country to avoid the natural price increase in its export goods. We can already see what that implies: wage reductions. Once again the relatively immobile factor of production is forced to pay for higher environmental costs.

Now you can see the paradox inherent in high-minded and well-intentioned efforts to compel a poor country to adopt local environmental standards that seem healthy and appropriate to those of us who live in nice, clean, rich countries. Improving the environment entails costs. Someone must pay those costs. I have tried to show that it is very likely that the costs will fall mainly on workers in poor countries. The Friends of the Earth may turn out to be the enemies of low-wage workers in Asia or Latin America. It may be too strong to think of that conclusion as a paradox, really. It is paradoxical only if we think we can ignore simple economic principles.

Does that conclusion mean that there is no case for environmentally friendly policies in the developing part of the world? No. I think that would not be the right inference. In principle, we could reason in the following way, though it may not be practical to do so.

For each imaginable level of world spending on pollution control or environmental improvement, one could try to list the best, most effective uses of that amount of resources, anywhere in the world. Doing so is not a purely technical matter; there would be powerful political forces at work in the choice between

cleaning up Lake Baikal or improving beaches on the Riviera. That is why I said that this is not a practical option; but it will help to clarify one's understanding of the problem.

If the world could decide just how much it wanted to spend on improving the environment, and if it had a list of environmental improvement projects, in descending order or urgency, it would know how it should spend that amount. Probably many of the desirable projects would be in poor countries. For example, improving air quality in Mexico City or reversing deforestation in the Amazon would have high priority on anybody's list of environmental projects around the world. Almost certainly, more of the desirable projects would be in poor countries than the poor countries themselves could pay for without reducing the rest of their standard of living by an intolerable amount. But the costs could be allocated separately through the device of an international environmental tax, with rich countries paying more than poor countries. The costs of local environmental improvement do not have to be paid by the direct beneficiaries.

I am not the village idiot. Of course, no such scheme could be carried out in practice. We have trouble enough in our own country in deciding which public expenditures to undertake and what kind of tax system should determine who pays for them. It would be far more difficult to get a rational, disinterested discussion at an international level. This fantasy is intended to make a much simpler and more general point. If rich countries are offended by the amount of environmental degradation in poor countries, rich countries should be prepared to subsidize pollution regulation in poor countries. Subsidies could be in monetary form, in the form of direct participation in pollution-abatement projects, or in the form of transfer of technology and direct technical assistance. The important thing is that catering to the sensitivities of the rich should not be at the expense of the poor. The use of trade restrictions for this purpose strikes me as especially inappropriate, or even immoral, because it serves to hide from ourselves exactly what we are doing.

This case seems to me to be fairly clear. We need only to understand what is going on to see what policy is right and what policy is wrong. There are other analogous situations that arise in current debates. They are similar in general structure but not identical in substance to the environmental case. I want to discuss briefly one or two of them here. The discussion has to be short and somewhat superficial; there is not time for more and I am less well equipped to have a strong opinion.

The case of child labor is an excellent example of the sort of issue I have in mind. In many of the poorest countries of the world, as we all know, large numbers of very young children are forced to work for low wages. The International Labor Organization has estimated that at least 120 million children under the age of 15 did full-time paid work in 1995. Most of them were between 10 and 14 years old, but millions were younger than that.

Those figures are inherently uncertain, partly because the definition of child labor is fuzzy, partly because measurement is difficult for any definition. But no one doubts that child labor is a serious matter in the Third World. Even apart from those who help their parents and older siblings in hard agricultural labor, many children are forced to give up education and the normal activities of childhood in order to earn money in a small factory or workshop setting. Some of them are, in effect, sold or rented by their parents to local entrepreneurs. I saw something like this in North Africa a few years ago, and it is unforgettably sad. I tried but could not force myself to imagine what it would be like if those had been my children or grandchildren.

One possible reaction of a European or American, perhaps the natural reaction, is to think, "We cannot allow this to happen. At a minimum, we must not participate in or benefit from the economy of child labor. If we do, we are guilty of encouraging it." The next practical step, of course, is to erect trade barriers against those goods that are produced with child labor or perhaps against all the exports of countries that engage in child labor. That is how this quite different conflict fits into the framework of this chapter.

There are important differences between pollution and child labor. Our initial feeling about child labor is that it is simply morally wrong to deprive children of education and the sorts of pleasures and freedoms that we normally associate with childhood. Some of us also think about environmental pollution primarily in moral terms but probably not nearly so intensely. We can easily imagine a society completely without child labor. We live in societies like that. We realize, on the other hand, that a society completely without pollution is not a realistic possibility; even those beautiful, pampered Bavarian cows emit methane. Pollution is inevitably a matter of degree.

Well, why does this matter? I have been emphasizing that an intelligent attitude toward "environmental dumping" must rest on a calculation of the consequences, all the consequences, of a proposed policy. In the case of trade restrictions, I have tried to show that the consequences are in some ways worse than the damage we are trying to avert. In the case of child labor, I can imagine someone saying that the consequences are irrelevant. Child labor is so contrary to simple humanity that any blow against it is justified, no matter what the side effects.

But suppose the side effects do harm to the very child we are trying to help? Then the consequences are certainly relevant. I can understand someone who says, "I will defend the principle of freedom of the press, even if some vulnerable and innocent people are harmed as a result." But it would be incoherent to say, "Child labor is morally wrong because it oppresses children. So I will not tolerate it, even if the alternative is worse for children." That just doesn't make sense.

Here I should mention that I owe whatever education I have about this problem to the work of Kaushik Basu, who is now Professor of Economics at Cornell University, and a consultant to the World Bank on questions of economic development.

The first thing to understand is that parents send their children to work for economic reasons, not because they are bad people. From the time of the Industrial Revolution in England to the Third World today, it has been observed that the children of better-off parents are less likely to work than the children of worse-off families. Unless you believe that earning power and moral goodness are very highly correlated, the obvious conclusion is that parents dislike sending their children out to work. They do so only when they are so poor that a child's earnings will make the family better off.

This has an interesting consequence. It is possible—by no means certain, but possible—that a complete prohibition of child labor could make labor so scarce that adult wages would rise to the point that parents would no longer want to send children out to work. Once that happens, the prohibition would no longer have any effect and could be removed. Professor Basu thinks that this kind of outcome might conceivably hold in richer Third World countries, but is very unlikely to hold in the poorer countries, where, in fact, most of the child labor is. Then a prohibition would make the poorest families worse off than they were before or else would drive children to work in uglier hidden or illegal activities (like prostitution), and the poorest families would still be worse off than they were before.

The case of total prohibition is not the immediately relevant one. Suppose, instead, that the advanced countries refuse to admit imports from industries that employ children. (Such legislation has, in fact, been proposed in the US Senate.) Then it is clear that children would merely be driven from employment in Third World export industries into worse jobs at lower wages in domestic industries. They would go from jobs in garment factories to harder and harsher jobs like breaking bricks or, once again, prostitution. The result would be to make both the children and their families worse off. It is a high price to pay for a feeling of self-satisfaction; but there are people who would choose that option.

We tend to think that this is primarily an issue between poor countries and rich countries. But developing countries are often in competition amongst themselves. Any one of them fears that its own efforts to abolish child labor will cause a shift in production to some other poor country with less stringent or less well-enforced laws.

The case of child labor is only the most dramatic of several proposals that have been made to use barriers to international trade as a way of imposing labor standards on (mostly) poor countries. This is more complicated than environmental standards, because there are direct benefits to workers to be offset against the incidental harm that occurs in the environmental example. I think that the only legitimate course is to analyze each such proposal one by one, always searching for the full consequences, not merely the intended ones. I do not think that one should jump to the conclusion that all measures of this kind are self-defeating; but skepticism is certainly a good place to start.

Conservatives sometimes use this law of unintended consequences as a reason to do nothing. I do not think it is a good reason, and I think that conservative tactic is wrong. In a changing world, doing nothing also has unintended consequences. I think the correct solution is to work harder at understanding the consequences of proposed policy actions. The poet and classicist A. E. Housman once said to his colleagues about a bad university decision, “Thirty seconds of thought would have resulted in a better decision. But thought is irksome, and thirty seconds is a long time.” I think he was joking.

This page intentionally left blank

PART IV

SOCIETY, POLITICS AND HISTORY

This page intentionally left blank

CHAPTER 22

PONDERING POVERTY, FIGHTING FAMINES

TOWARDS A NEW HISTORY OF ECONOMIC IDEAS

SUGATA BOSE

In the history of economic ideas in India the specter of famine looms large in conceptions of poverty. The two terms came to be paired in the title of Amartya Sen's 1981 classic, *Poverty and Famines: An Essay on Entitlement and Deprivation*. Sen was careful in delineating the connection between the two while spelling out what distinguished poverty from starvation as well as starvation from famines. Sen's entitlement approach is pertinent both to the analysis of famines in history and a re-evaluation of the history of economic ideas that have developed at the intersection of poverty and famines.

The credit for turning poverty and famines into a subject of history and political economy rather than a feature of the ancient and natural order of things must go to Indian economic thinkers of the late nineteenth and very early twentieth

century. They not only managed to engage and defeat their colonial masters in a sophisticated intellectual debate, but also disseminated the changing terms of the discourse to a broad cross-section of politically aware society. The pioneering figure in this respect was the indefatigable Dadabhai Naoroji, who began writing on poverty against the backdrop of the 1870s famines (Naoroji 1901/1962). But it was the public discussion surrounding the devastating famines of 1897 and 1900 that took Indian ideas about poverty to a new level of reasoned, if impassioned, articulation.

"The shadows darkened and deepened in their horrors as the year advanced," Mahadev Govind Ranade grimly recorded in 1897, "and it almost seemed as if the seven plagues which afflicted the land of the Pharaohs in old time were let loose upon us, for there is not a single province which had not its ghastly record of death and ruin to mark this period as the most calamitous year of the century within the memory of many generations past" (Ranade 1992: 180). This was the year in which Britain was celebrating the diamond jubilee of Queen Victoria's reign with much pomp and ceremony. It was also the twentieth anniversary of her assumption of the title "Empress of India", which had been translated for the benefit of her colonial subjects as "Kaiser-e-Hind". Romesh Chunder Dutt opened his 1903 book, *The Economic History of India in the Victorian Age*, with a reference to "a celebration in London" six years ago that was "a scenic representation of the Unity of the British Empire". Indian princes rubbed shoulders at this gathering with "loyal Canadians and hardy Australians". "One painful thought, however," Dutt pointed out, "disturbed the minds of the people. Amidst signs of progress and prosperity from all parts of the Empire, India alone presented a scene of poverty and distress. A famine, the most intense and the most widely extended yet known, desolated the country in 1897" (Dutt 1903/1969: p. v). Dutt may have been off the mark in his claim about Indian exceptionalism, since the sun never set on "late Victorian holocausts" (Davis 2001). But he was right about the scale of the calamity that had tarnished the jewel in the Victorian crown. After a brief respite in 1899 another huge famine gripped India from 1900. It lasted nearly three years and took 1.25 million lives according to official figures, and close to five million in the estimates of Indian nationalists. When the Viceroy of India, George Nathaniel Curzon, organized his glittering Delhi Durbar in January 1903, there were still tens of thousands of famine victims enduring a wretched existence in relief camps.

Curzon was the most eloquent early twentieth-century exponent of the food availability decline (FAD) approach to famine, which Amartya Sen eventually laid to rest in 1981. In 1900 the Viceroy squarely held India's capricious monsoons and the consequent reduction of agricultural output to be responsible for the series of 1890s famines. According to Curzon, no administration could "anticipate the consequences of a visitation of nature on so gigantic and ruinous a scale". In 1902 he claimed that "to ask any Government to prevent the occurrence of famine in a country, the meteorological conditions of which are what they are here and the population of which is growing at its present rate, is to ask us to wrest the keys of

the universe from the hands of the Almighty" (quoted in Chandra 1966: 49). This was a valiant attempt to pass off a man-made catastrophe as an act of God, and a striking case of an obsession with FAD leading to the failure of policies aimed at averting famines. R. C. Dutt resoundingly rejected the FAD explanation of famines. As he put it succinctly:

> The food supply of India, as a whole, has never failed. Enough food was grown in India, even in 1897 and 1900, to feed the entire population. But the people are so resourceless, so absolutely without any savings, that when crops fail within any one area, they are unable to buy food from neighbouring provinces rich in harvests. The failure of rains destroys crops in particular areas; it is the poverty of the people which brings on severe famines.
>
> (Dutt 1903/1969: p. vi)

However terrible the scourge of the famines and the high mortality they brought in their trail, Dutt was convinced that they were "only an indication of a greater evil—the permanent poverty of the Indian population in ordinary years" (Dutt 1903/1969, p. vi). Naoroji similarly believed that India was "able to grow any quantity of food", but "a large proportion of the population was in a normal state of starvation" (Naoroji 1901/1962: 579). This emphasis by the doyens of turn-of-the-century Indian economic thought on shortage of purchasing power rather than shortage of food may be seen, to borrow Sen's formulation, as "a rudimentary way of trying to catch the essence of the entitlement approach" without its "comprehensive account of a person's ability to command commodities in general and food in particular" (Sen 1981: 155–6).

Curzon's other Malthusian fallacy—holding rapid population increase and over-population to be a cause of poverty and famines—had been anticipated and rebutted by another astute economic thinker, G. V. Joshi, more than a decade earlier. In an 1890 essay he had started from the premise that "increase of numbers is *per se* not necessarily or always an evil, as Malthusian writers assert". He followed it up by comparing India's demographic and developmental trajectory with the role of population increases and densities in Western Europe's economic progress to argue that for a country like India a rise in population may well be a boon rather than "a curse as Malthusian economists would expect". He also disposed of yet another colonial explanation for India's economic misery—the improvidence of the Indian peasantry. "If our earning power is so low as at present and our income hardly ever comes up to the level of our necessary expenditure," Joshi wrote, "the evil does not lie in our over-spending propensities, but in those conditions of industrial life in this country which keep our earning so low." A comparative glance at the French peasantry after the end of feudalism was sufficient to show that the Indian predicament did not flow from its peasantry's bad habits but from the state's bad economic policies (Joshi 1912: 773, 775).

The debates about India's growing poverty that had been raging during the last three decades of the nineteenth century became even more topical during the great

famines of 1897 and 1900. Figures on India's abysmally low per capita income compared with other countries, including Britain, and estimates of basic subsistence needs, as revealed in expenditure on prisoners, did not remain confined to academic disputes once the colony was in the throes of massive famines. The biological definitions of poverty that had been in vogue in the late nineteenth century, to borrow an insight from Sen once more, now seemed in need of reformulation and not absolute rejection (Sen 1981). A premier British medical journal, *The Lancet*, estimated famine mortality in India during the 1890s at 19 million, which, William Digby underscored, was about half the population of Britain (Digby 1901). The death toll for the 1897 famine ranged from the official figure of 4.5 million to unofficial claims of up to 16 million (Davis 2001: 158). Economic ideas were now being traded in a highly charged political environment.

In addition to trotting out the FAD explanation, beleaguered colonial officials took refuge behind claims that India had suffered poverty since time immemorial, and that novel bureaucratic arrangements had been put in place by the colonial state since the 1870s to provide famine relief. Nineteenth-century colonial rhetoric had already painted the eighteenth century as a dark age of chaos and anarchy that had been replaced by the golden age ushered in by the *Pax Britannica*. Given the strength of this discourse, hardly anyone—not even the government's staunchest critics—thought to point out that the first seventy years of the eighteenth century had been remarkably free of large-scale famines and was, in fact, a period of general agricultural prosperity. The first major eighteenth-century famine occurred in 1770, close on the heels of the East India Company's assumption of the *diwani*, the right to collect the revenues of Bengal. Somewhere between a fifth and a third of Bengal's population perished in that famine. This early colonial catastrophe was no Malthusian event; the state and the economically powerful had combined to create declines in entitlements to food of vulnerable social groups, turning, in Adam Smith's phrase, a "dearth into a famine" (Bose 1993: 17–24).

Even though Indian economic thinkers in the early twentieth century did not wage their intellectual battles on the contested ground of the eighteenth century, they did come forth with an alternative narrative of history to question the colonial autocracy's claims that India's poverty went back to antiquity. Their argument was that the layered and shared sovereignty of India's pre-colonial past had been better suited to serve the country's economic needs than a centralized and unitary British administration. R. C. Dutt stated the position with his characteristic clarity:

> Such was not the past in India. Hindu and Mahomedan rulers were always absolute kings, often despotic, but never exclusive. Their administration was crude and old-fashioned, but was based on the co-operation of the people. The Emperor ruled at Delhi; his Governors ruled provinces; Zemindars, Polygars, and Sardars virtually ruled their estates; villagers ruled their Village Communities. The entire population, from the cultivator upwards, had a share in the administration of the country. (Dutt 1903/1969: 615)

What gave Indian critics of the government a decisive advantage over its defenders was their insistence on a public inquiry and debate. The Secretary of State for India had resisted the demand for such an inquiry in the aftermath of the 1897 and 1900 famines. The last official inquiry into the economic condition of the poorer classes of India had been conducted during the viceroyalty of Dufferin in 1888. Its results were kept a closely guarded state secret. R. C. Dutt, along with many other of his compatriots, berated the government for the "concealment of facts". The poverty of the Indian people was, after all, "not a secret" and an evil could hardly be remedied "by being hidden from the eye" (Dutt 1903/1969: 608). The economic ideas of Indian thinkers were communicated to a broader audience through a network of newspapers and an expanding press and publications market. Colonial attempts to curb freedom of expression generally proved counter-productive and curtailed the government's own ability to fight famines. One example of this tendency was the decision to charge Balwantrao Gangadhar Tilak for sedition when his Marathi paper, *Kesari*, published some verses in 1896 titled "Shivaji's Utterances". "What desolation is this!", Shivaji chanted about the effects of British rule; "plenty has fled and after that health also. The wicked *Akabaya* [misfortune personified] stalks with famine throughout the whole country" (quoted in Chandra: 1966: 11). The government succeeded in convicting Tilak, but failed miserably to mitigate even slightly the dire consequences of the 1897 famine, which brought utter ruin to the cotton-growing tracts of western India.

The government attempted to gag the press in India, where famine stalked the land, but the press was free in Britain. So newspaper readers in London and Manchester were well aware of the tragedies unfolding in India during the 1897 and 1900 famines. The best that concerned citizens in the imperial metropolis could do for their less fortunate fellow-subjects of Queen Victoria was to contribute to charity in aid of the relief work in India. The charitable cause of Indian famines had to battle for and usually cede newspaper space to other imperial events, such as the Boer War in South Africa. For Indian critics, charity was not the answer to the problem of poverty and famines. This had been caused by colonial policies and could only be remedied by changing them. R. C. Dutt summarized his remedy in two words: "retrenchment" and "representation". Under the category of retrenchment he included, most importantly, "a reduction in the imposts on land", but also a repeal of the excise duty on Indian cotton manufactures, a cut in Indian military expenditure, and a greater Indianization of the administration to lower the remittance of Home Charges, which included pensions for British officials who had served in India. By "representation" he meant a moderate dose of the elective principle in governing India. "Is it any wonder", he asked, "that ... the oligarchy at Whitehall and the oligarchy at Simla should, amidst surrounding Imperial influences, sometimes forget the over-taxed Indian cultivator, the unemployed Indian manufacturer, the starving Indian labourer?" (Dutt 1903/1969: 612–15).

Indian economic thinkers won their victory in the conceptual domain. They shifted the discourse on poverty and famines from the sphere of nature to the arena of political economy. In their indictment of British rule, these critics were not always on target when it came to the precise colonial policies they believed culpable for the economic ills afflicting their country. R. C. Dutt, for example, blamed the spate of late nineteenth-century famines on the high land revenue demand of the colonial state. Deploying the method of cross-regional comparison, he was able to show that provinces shouldering the highest burden of the land tax had the highest famine mortality rates. This attempt to establish proof by correlation fell short of the exacting standards required to establish a clear chain of causation. Moreover, the revenue and rent offensive that had begun at the onset of the nineteenth century had been waning in most parts of India from the 1870s and 1880s. Tenancy legislation in the 1880s had assured a degree of security of tenure and a moderation in rents in the permanently settled areas, where famine mortality was the lowest. However, more generally, debt interest had replaced revenue and rent as the principal means of appropriating the surplus of the peasantry. Financial policy rather than fiscal policy alone—that is, colonial initiatives in the spheres of currency and credit, rather than just the heavy hand of the colonial land revenue burden—would have been a more plausible target for critics of the government. Indian intellectuals speaking in the name of an Indian nation in the making, however, could not help tailoring their discourse to their global and metropolitan audience, and to the encompassing multi-class constituency of which they saw themselves as spokesmen.

Another tier of economic ideas tended to be more responsive to recent shifts in the mechanisms of exploitation: ideas expressed through proverbs and folk-poems that belonged to what is sometimes regarded as an intermediate, if not quite lowbrow, level of economic thought. "I love the Sheth-Vaniya so much", went one Rajasthani Bhil proverb, "that I have given him a fat stomach" (Hardiman 1987: 3). A rustic poet of Bengal, Maulavi Idris Ali, writing in a small tract published in 1921, had grasped the source of the indebted peasant's descent into poverty:

> *Jedin theke mahajaner shange janajani*
> *Shedin thekei Kalu Sheikher bejay tanatani.*
>
> (Ever since he made the *mahajan*'s acquaintance
> Kalu Sheikh has been in straitened circumstances.)

The creditor, unlike the rentier, played the role from the late nineteenth century of giving and taking, nurturing and impoverishing, at the same time. One popular folk-poem after another in the early twentieth century recorded and lamented how debt followed in the wake of increased need for credit, rather than greater creditworthiness flowing from prosperity, during the decades of rapidly expanding commodity production in rural India for the world market (Bose 1986: 80–1; 1993: 122–3).

The historiography on economic ideas in India in the first four decades after independence annexed the work of key thinkers a little too easily to a thesis about "economic nationalism" (Chandra 1966) that sustained post-colonial India's inward-looking policies during those decades. While leaving the "lower" domains of thought out of consideration, such an approach did scant justice to the "elite" thinkers' attentiveness to regional differentiation within India as well as their transnational and global vision that transcended the borders of India. This vision included engagement with theory crafted in other regional contexts, such as Germany and Central Europe, and empirical comparisons with countries covering a very wide geographical spectrum. A new scholarly agenda on the transnational history of political and economic ideas is still in its infancy (Bayly 2007). Some preliminary work has begun to appear on political ideas which may liberate the study of economic ideas from its Swadeshi mode. The normative ideas and political practice of the economic thinkers who offered a sharp indictment of colonial policies used to be seen as leading to a political movement, beginning in 1905, that put the accent on the protection of indigenous identities and institutions. That movement is now being interpreted in a new transnational frame. The Swadeshi milieu, despite its interest in rejuvenating home-grown traditions and industries, was not wholly inward-looking in its cultural and economic orientation. The protagonists of Swadeshi were curious about innovations in other parts of the globe and they certainly felt comfortable within ever-widening concentric circles of Bengali patriotism, Indian nationalism, and Asian universalism. After the government cracked down on them in 1908, Swadeshi figures literally wandered beyond the confines of their home into the wider world (Manjapra 2007). The challenge facing political and economic thinkers of the Swadeshi era was how to reconcile their sense of Indian nationality with an aspiration towards a common humanity. In this endeavor hardly any of them allowed colonial borders to constrict their economic and political imagination.

Yet later anti-colonial thinkers differed in their interpretation of the economic tomes of turn-of-the-century critics of colonial policy. "When I read Mr Dutt's economic history of India," Gandhi wrote in *Hind Swaraj* in 1908, "I wept; and as I think of it again my heart sickens. It is machinery that has impoverished India" (Gandhi 1958: 22–3). While Gandhi condemned the evils of modern industrialism as the cause of India's poverty, many of his contemporaries denounced colonial fiscal and financial policy for having negated India's potential for industrialization, which was seen as vital to reducing poverty. Even at the height of the Indian famine of 1900, Dadabhai Naoroji had responded to a suggestion during a debate in London that in order to combat famines India "should look more to manufacturing industries and be less dependent upon agriculture". Naoroji stated emphatically that India was "originally noted for her industries" and that "the Indian industries had been destroyed by the British policy" (Naoroji 1901/1962: 582). Three decades after Gandhi's *Hind Swaraj* the President of the Indian National Congress, Subhas

Chandra Bose, reckoned in 1938 that the eradication of poverty was the principal problem of reconstruction facing the country. It would require radical land reforms, including "the abolition of landlordism" as well as the liquidation of agricultural indebtedness and the provision of "cheap credit to the rural population". But to solve the problem of poverty, agricultural improvement "would not be enough", and a plan of state-led industrial development would be necessary. "However much we may dislike modern industrialism and condemn the evils which follow in its train," Bose declared, "we cannot go back to the pre-industrial era, even if we desire to do so" (quoted in Bose 1997: 49). The National Planning Committee that he formed contained as one of its members a Gandhian, J. C. Kumarappa, who forcefully articulated the argument against industrialization and Meghnad Saha, the eminent scientist, who believed in exactly the opposite. The Committee, chaired by Jawaharlal Nehru, served as a venue for lively economic debates animated by a common ambition of reducing and ultimately removing poverty.

Four points need emphasizing about the ideas of development discussed by Indian thinkers in the inter-war period. First, the discourse on national development was not parasitic on colonial development, an idea that did not really jell until the British parliament passed the Colonial Development Act in 1944. Second, what was "national" in ideas of development did not prevent an engagement with theories and practices well beyond the "nation". Third, the emphasis in this period was on the idioms rather than merely the instruments of development, or what Amartya Sen later described as "means use" as distinct from "means enhancement" in thinking about development (Sen 1983). Fourth, economic ideas were articulated in tandem with political ideas that displayed great variation in preferences regarding the spirit and form of the state that ought to replace the British raj in India. Acquisition of power in a unitary, centralized state was by no means a foregone conclusion. It was a topic on which even self-avowed socialists espoused different opinions, as became evident in Meghnad Saha's powerful critiques of Nehru's over-centralizing policies in the immediate post-colonial era (Chatterjee 1987). These four elements might serve as useful guiding principles in any new exploration of the history of economic ideas. They would also enable scholars to avoid the pitfalls of viewing development as necessarily a project of oppressive, centralized states, as is the wont of votaries of "anti-development", and keep in view a full consideration of the complex intellectual history of "development as freedom" (Sen 1999).

While Indian ideas about poverty and development in the early twentieth century were nourished by a strong argumentative intellectual tradition, colonial ideas about famines showed a remarkable insensitivity to the dramatically changing political and economic context and the flow of new ideas. As Amartya Sen showed in his 1981 book, an obsession with the FAD explanation for famines contributed in no uncertain way to the colonial state's abject failure on all fronts during the great Bengal famine of 1943. The famine, which began in March 1943, was not acknowledged in the British parliament until October of that year. The fabled "famine code" of the nineteenth century was not even invoked by the colonial state.

External supplies of food for famine relief were either not sought or not accepted. As Sen noted, even if a famine is not caused by decline in food availability, large injections of food into the public distribution system are needed to break its grip on the economy and society (Sen 1981: 79).

The British rulers of India were not the only ones culpable for large-scale famine fatalities during the Second World War. The Guomindang regime in China, backed by the United States, and the Vichy French regime in Vietnam, propped up by the Japanese, played an equally dismal and dishonorable role. Three gigantic wartime famines devastated Bengal in India, Honan in China, and Tonkin in Vietnam; in each of these somewhere between 2 and 3 million people died, all of them cases of "starvation amidst plenty" (Bose 1990). All three were "boom famines" that took their hefty toll in human life and suffering during the inflationary conditions of the war. Vulnerable social groups and families lost their entitlement to food in a period of soaring grain prices. The scale of the impact of the price and market mechanism on entitlement relations cannot, however, be understood without a reference to the global economic context. Bengal, Honan and Tonkin were old, complex agrarian societies, characterized until recently by diverse forms of appropriation, distribution and exchange. The almost exclusive dependence on entitlement to food in volatile product markets had much to do with the rupturing of credit relations during the worldwide depression of the 1930s. These ties had underpinned an uneasy and unequal symbiosis in social relations from the late nineteenth century to 1930. In the absence of credit, the only way to obtain money for food was to sell movable property and land, which in turn entailed a long-term loss of direct entitlement to food.

In his analysis of the Bengal famine Amartya Sen tracked the steep decline in wage–price exchange rates between 1939 and 1943 against wage labor. The greatest mortality and destitution was therefore to be found among the occupational category "landless agricultural labor". Sen also discussed the decline in exchange entitlements to rice among those who produced other commodities, such as wheat, mustard oil, cloth, bamboo umbrellas, milk, fish and haircuts (Sen 1981: 67–70). The smallholding jute cultivators were critically dependent on relative prices and especially vulnerable to that determinant of entitlement. The famine was at its most intense in the jute-growing districts of east Bengal (Bose 1986: 94–5).

An emphasis on the "exchange entitlement" aspect of Sen's theory ought not to obscure the finer points he made on the subject of "direct entitlements". These are quite pertinent to his own analysis of the 1970s famine in the Wollo region of Ethiopia and can help explain why the Bengal famine affected a whole spectrum of the rural poor, including rice-growers. To use Sen's words, "the immediate influence affecting starvation is the decline in the food grown and owned by the family, rather than a fall in the total output in the region as a whole" (Sen 1981: 101), a key distinction, since the FAD approach is based on the latter. In certain parts of Bengal in 1943 many would-be net sellers had turned into net buyers of rice for local reasons and would-be small buyers became larger buyers; some had sold early and been

caught out by the subsequent upswing in prices. The failure of direct entitlement can compel a peasant to rely on the food market at a time when his exchange entitlement has been damaged. The terrible predicament in the Bengal famine of smallholding rice cultivators, not just laborers and jute-growers, indicates the need to examine how fluctuations in food-grain availability for peasant family economies relate to their exchange entitlement in the food market, and the implications of that relationship for starvation and famine. Sen's insights also make it imperative to investigate the divergences in entitlements and capabilities within families along lines of gender and generation. Most famines are "man-made", both in the sense of human agency as opposed to fate or divine ordinance, and specifically male decision-making in the social selection of famine victims. The power and elegance of Sen's "entitlements and capabilities approach" lies in its ability to address the intricate nuances of specific famines while simultaneously advancing a general theory of poverty.

When Indian economic thinkers engaged in the debate about poverty with their colonial masters in the late nineteenth and early twentieth century, famines were typically cited as the dramatic demonstration, if not conclusive proof, of India's deepening poverty. This close link between poverty and famines in the history of Indian economic thought may have had an unintended negative consequence in post-independence India. With the banishment of large-scale famines from democratic India, persistent poverty became less spectacular in its manifestation. It was once more left to Amartya Sen to point out that chronic hunger and malnutrition did not appear to upset India's democratic system, even though mortality rates higher than China's in normal years (rather than the years of the catastrophic Chinese famine of 1959–61) ought to have been grounds for grave disquiet (Sen 1982).

The best tribute that historians of South Asia can pay to its greatest economist, philosopher and man of ideas is to launch an innovative scholarly agenda for investigating anew the history of economic ideas. It would mean traveling an intellectual path that crosses the borders of the disciplines of history and economics and locating South Asian thinkers along the "arcs of ideas" (to borrow Emma Rothschild's felicitous phrase) that traverse a transnational trajectory (Rothschild 2006). Placing South Asia in the vortex of the cosmopolitan thought-zones of the globe would be an endeavor worthy of one of the finest cosmopolitan intellectuals of the modern age.

References

Bayly, C. A. (2007), "Afterword", *Modern Intellectual History*, 4: 163–9.

Bose, Sugata (1986), *Agrarian Bengal: Economy, Social Structure and Politics, 1919–1947* (Cambridge: Cambridge University Press).

—— (1990), "Starvation amidst Plenty: The Making of Famine in Bengal, Honan and Tonkin, 1942–45", *Modern Asian Studies*, 24: 699–727.

—— (1993), *The New Cambridge History of India: Peasant Labour and Colonial Capital* (Cambridge: Cambridge University Press).

—— (1997), "Instruments and Idioms of Colonial and National Development: India's Historical Experience in Comparative Perspective", in Frederick Cooper and Randall Packard (eds), *International Development and the Social Sciences: Essays on the History and Politics of Knowledge* (Berkeley, Calif.: University of California Press).

Chandra, Bipan (1966), *The Rise and Growth of Economic Nationalism in India: Economic Policies of Indian National Leadership, 1880–1905* (New Delhi: People's Publishing House).

Chatterjee, Santimoy (ed.) (1987), *Collected Works of Meghnad Saha*, vol. 2 (Calcutta: Saha Institute of Nuclear Physics).

Curzon, George Nathaniel (1904), *Speeches*, vol. 3 (Calcutta: Office of the Superintendent of Government Printing).

Davis, Mike (2001), *Late Victorian Holocausts: El Niño Famines and the Making of the Third World* (New York: Verso).

Digby, William (1901), "*Prosperous*" *British India: A Revelation from Official Records* (London: T. F. Unwin).

Dutt, Romesh (1903/1969), *The Economic History of India in the Victorian Age: From the Accession of Queen Victoria in 1837 to the Commencement of the Twentieth Century*, vol. 2 (New York: Augustus M. Kelley).

Gandhi, Mohandas Karamchand (1958), *The Collected Works of Mahatma Gandhi* (New Delhi: Publications Division, Government of India).

Hardiman, David (1987), "The Bhils and Shahukars of Eastern Gujarat", in Ranajit Guha (ed.), *Subaltern Studies V: Writings on South Asian History and Society* (Delhi: Oxford University Press).

Joshi, G. V. (1912), *Writings and Speeches* (Poona: Arya Bhushan Press).

Manjapra, Kris Krishnan (2007), "The Mirrored World: Cosmopolitan Encounter between Indian Anti-Colonial Intellectuals and German Radicals, 1905–1939", unpubl. Ph.D. diss., Harvard University.

Naoroji, Dadabhai (1901/1962), *Poverty and Un-British Rule in India* (New Delhi: Publications Division, Government of India).

Ranade, Ramabai (ed.) (1992), *Miscellaneous Writings of the late Hon'ble Mr. Justice M.G. Ranade* (Delhi: Sahitya Akademi).

Rothschild, Emma (2006), "Arcs of Ideas: International History and Intellectual History", in Gunilla Budde, Sebastian Conrad and Oliver Janz (eds), *Transnationale Geschichte: Themen, Tendenzen und Theoren* (Gottingen: Vandenhoeck and Ruprecht).

Sen, Amartya (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Clarendon Press).

—— (1982), "How is India Doing?", *New York Review of Books*, 16 December.

—— (1983), "Development: Which Way Now?", *Economic Journal*, 93: 745–62.

—— (1999), *Development as Freedom* (Oxford: Oxford University Press).

CHAPTER 23

IDENTITY, VIOLENCE AND THE POWER OF ILLUSION

JONATHAN GLOVER

It is a commonplace that Amartya Sen has played a huge role in broadening the conceptual base of economic theory. Central has been his insistence that explaining what people do in terms of preferences is much too simple. One simplification is the failure to distinguish between different motives, some of which are not directed towards personal gain. People can act out of sympathy for others or out of commitments—perhaps moral, political or religious—that may conflict with narrow self-interest. He persuasively argues that a single ranking should be replaced by a "ranking of rankings".

Sen's discussion of commitments makes it clear that their core does not have to be either personal or universal: "Groups intermediate between oneself and all, such as class and community, provide the focus of many actions involving commitment" (Sen 1977: 109). He also signals awareness that these group commitments can be bound up with issues of identity: " 'We' demand things; 'our' actions reflect 'our' concerns; 'we' protest at injustice done to 'us'. This is, of course, the language of social intercourse and of politics; but it is difficult to believe that it represents nothing other than a verbal form, and in particular no sense of identity" (Sen 2002: 215).

Characteristically, Sen's attunement to the role of identity in these group commitments is combined with arguments showing that much of the resulting identity politics rests on simplification and illusion. In particular he argues against two "unfounded assumptions". The first is the illusion of the simplicity of identity: the belief that "we must have a single—or at least a principal and dominant—identity". The second is the illusion of destiny: the belief that "we 'discover' our identity with no room for any choice" (Sen 2005: 350).

Sen points out that the illusion of the simplicity of identity depends on points we overlook. These points, when stated, are obvious truths, but the truisms are given an edge by an example that testifies to the scale of the harm the illusion can bring about: "A Hutu from Kigali may be pressured to see himself only as a Hutu and incited to kill Tutsis, and yet he is not only a Hutu, but also a Kigalian, a Rwandan, an African, a laborer, and a human being." This is one of many cases where so much horror would have been avoided if the illusion had been replaced by what Sen calls "the recognition of the plurality of our identities and their diverse implications" (Sen 2006: 4).

For Sen it is also important to challenge the illusion of destiny: "there is a critically important need to see the role of *choice* in determining the cogency and relevance of particular identities which are inescapably diverse" (2006: 4). The illusion that, among the plurality of our identifying features, those that constitute the core of our identity are destined rather than chosen has the effect of "impoverishing the power and reach of our social and political reasoning". Through this it "exacts a remarkably heavy price" (2006: 17). A reminder of some aspects of the price is the justified protest of Richard Dawkins that, since young children cannot decide for themselves about religious beliefs, there is no such thing as a Christian child or an Islamic child. There are only children of Christian or Islamic parents (Dawkins 2006: 379–83) Part of this price is the stifling of people's freedom to think for themselves and to create themselves. Part of it is seen in the tribal conflicts, both national and religious, that kill so many people round the world.

I agree with Amartya Sen both that these beliefs about the simplicity and inevitability of identity are illusions, and that they do enormous harm. So my aim in this chapter is only marginally critical and mainly constructive. I hope to add to the discussion of why these illusions have such power, in the hope of weakening their grip. The approach will be by way of asking two questions. First, why are certain of people's characteristics (ethnicity, nationality, language, religion, etc.) so often taken by people to be central to their identity, in a way that others (their job, height, taste in music, etc.) usually are not? What is the hold of *these* characteristics on our sense of who "we" are? And second, why is this hold on us so often linked to violence between "us" and "them"?

I. Identity and the Charged Characteristics

Why *these* characteristics? Why do nationality and religion, for example, figure so large in people's conception of their own identity? What gives being Christian or Islamic, Russian or Polish, the kind of emotional charge that makes a slight to the group seem an attack on a person's very core? Why is a sneering remark about Polish plumbers more likely to anger other Poles than other plumbers?

The conjecture I want to put forward is that part of the explanation of these charged characteristics touches on two very deep human needs. We want our life to add up to something. And there is the need for recognition by others.

We want our lives to add up to something, and whether or not they do so is partly up to us. Of course, much in our lives is beyond our control. We do not choose when we are born, whether we are born male or female, whether we are born with or without medical problems, who our parents are, or the culture in which we grow up. But, despite the huge constraints on what choices are open to us, as we grow up most of us are able increasingly to shape the way our lives turn out and the kinds of people we are. There are the cards we are dealt but there is also how we decide to play them.

John Rawls suggested that people have a life plan (Rawls 1971: 407–16) But perhaps only a few philosophers and other obsessives have anything so explicit. The related projects of shaping our lives and shaping ourselves are, for most of us, most of the time, not things we are much conscious of. But still, at the back of our minds we often have an awareness of a vaguely spelt out narrative of our life, and of how its general direction is partly guided by things we care about. Galen Strawson has rightly pointed out that in this respect people differ psychologically. Some people may not think of their lives in terms of a narrative and this may not make their lives less good (Strawson 2005). And others may be too ground down by poverty, war or oppression to give attention to what their lives add up to. But, despite all this, how most people react when their autonomy is under threat brings out the importance of our intermittently conscious power partly to shape our own lives.

The charged characteristics that figure so much in people's sense of their own identity are often those intimately bound up with the narratives of their lives. Characteristically, a narrative is bound up with places: where we grew up, where we were in love, where our parents died or where our children were born. Part of the pain of exile is being cut off from all this. The narrative is also likely to be permeated by the beliefs and values (whether religious, moral or political) that give point to some of what we do. Most narratives are bound up with (a few) other people: often family or close friends. What I did was done with them, or in response to what they said or did. My values were partly shaped by them. Perhaps I still carry their hopes

and expectations with me. The narrative we tell may sometimes partly have them as the imagined audience, to a degree we realize perhaps only when they die.

We often want recognition from other people. At the very minimum, we do not want others to look through us as if we were not there, as in the cruelty of ostracism. (The extreme of this cruelty is solitary confinement, where people are put away so they are not even seen.) The minimum is being noticed as a person, but we need more than this. We want people to see what we are like, what is distinctive about us. In this way the need for recognition is linked to caring that our lives add up to something. We want at least a few people close to us to notice how we are shaping our lives and ourselves. If this does not happen, our self-creation paints a picture no-one sees, or writes a novel no-one reads.

The list of charged characteristics typically includes the religious or other beliefs that have guided the choices in the narrative. It includes the places that have been important, and the language and culture shared by many who play a major part in it. A lot of the meaning of what we do derives from its context. So insults to the place where we grew up, or to our religion, or to our national or ethnic group, strike at much of what gives our personal narrative its sense.

Usually, many of the people close to us, on whose recognition we depend, have charged characteristics overlapping with ours. So insults to our nation, ethnicity or religion, or to our language or our country, also devalue the group whose responses validate our narrative.

If true, this conjecture, about the links between the charged characteristics and our deep need for meaning and for recognition, may partly explain why the hold on us of tribal loyalties is so strong. The grip of these loyalties is not just on those who are simple and unreflective. It can be seen in people who are fully aware of the harm done by the illusion of the simplicity of identity and by the illusion of destiny.

Few are more sophisticated about these illusions than Salman Rushdie. His family has Indian and Pakistani branches, and he has often found himself trying to erode the growing prejudices by explaining Indian or Pakistani attitudes to those on the other side. But he admits he has not always been entirely impartial: “I hate the way in which we, Indians and Pakistanis, have become each other’s others, each seeing the other as it were through a glass, darkly, each ascribing to the other the worst motives and the sneakiest natures. I hate it, but in the last analysis I’m on the Indian side” (Rushdie 2002: 431.)

If Salman Rushdie cannot escape the bias towards “his” side, it is unlikely that many of the rest of us will find it easy to do so. Amartya Sen’s use of rational criticism to undermine illusions about identity is the beneficial performance of a necessary task. But the case of Salman Rushdie suggests that, even when rational argument has shown us that we see each other through a glass darkly, the glass may not immediately become clear. Part of the explanation may be that our own need for meaning and recognition continues to cast a darkening shadow of tribal identity.

II. The Links Between Identity and Violence

What is it about the sense of tribal identity that leads to violence? The links depend on a complex interplay between our inner psychological dispositions and historical and social circumstances.

II.1 The Disposition to Group Hostility

Part of the inner psychology is a disposition towards group hostility that is independent of the charged characteristics and the sense of identity. It is a platitude that human beings are drawn to identify with the group they belong to, in ways linked to violence towards other groups. What is more surprising is that such violence sometimes breaks out where the groups are *not* marked off from each other by any ethnic, religious, or other charged characteristic.

Philip Zimbardo's famous 1971 study of students in a "prisoners and prison officers" role-playing experiment showed violence between "minimal groups": those whose differences are known to be marginal or non-existent. The study had to be called off after only a few days because the "guards" treated the "prisoners" with brutality, including sexual humiliations approaching those more recently in Abu Ghraib:

> It is hard to imagine that such sexual humiliation could happen in only five days, when the young men all know that this is a simulated prison experiment. Moreover, initially they all recognized that the "others" were also college students like themselves. Given that they were all randomly assigned to play these contrasting roles, there were no inherent differences between the two categories. (Zimbardo 2007: 172)

Evolutionary psychologists have speculated that, in the Stone Age environment of early human hunter-gatherer tribes, a genetic disposition towards hostility to other groups may have been conducive to survival (Shaw and Wong 1988). Whether or not this is right, the Zimbardo study of violence between minimal groups does suggest a disposition to what can be called "basic group hostility": a hostility to other groups that is independent of illusions about their characteristics and identity.

Obviously, not all contact between minimal groups elicits brutality and violence. A prison role-playing experiment is a very particular situation. One group is given nearly total power over the other. And the victim group is subjected to a dehumanizing anonymity of dress and behavior. Many ordinary situations do not bring out barbarism in the same way. Members of two different orchestras queueing up to leave coats in the cloakroom of a concert hall do not create a small Abu Ghraib. Their situation suppresses, or at least does not elicit, basic group hostility.

The disposition the Zimbardo study tapped into is more likely to turn into violent conflict when links are made to our identity and its associated charged characteristics. We would be surprised to turn up for a concert to find that the members of one orchestra had killed half the members of another. It is less surprising when religious differences create a crusade or a *jihad*, or when ethnic or national differences create a war of conquest or of liberation.

But, normally, group hostility is not aroused simply by encounters between groups who see their identity in terms of different nationalities and religions. No alarm signals are triggered when half a hotel is taken by a busload of Italian Catholics and the other half by a busload of Swedish Protestants. The smoldering disposition to group hostility, even when combined with charged identity differences, catches fire only in certain social and historical circumstances. Group narratives give a clue to which contexts make relations between groups more combustible.

II.2 Group Narratives

Indians, Russians or Greeks may think of themselves as each making up a distinct community. But, as Benedict Anderson has famously argued, nations are "imagined communities" (Anderson 1983). Russians are not a community in the way that people in a small village are. When the Russian nation is angry about some insult, this requires a degree of imaginative creation, usually helped along by television and newspapers. Nations are not the only imagined communities. The same goes for Christendom, the Islamic *umma*, the Greek Orthodox Church, and so on. Communities are created partly by actual social construction, such as the passing of laws, the use of a common currency, or the adoption of creeds. They are created partly by physical construction, such as the making of a transport system or the building of cathedrals, synagogues or mosques. They are created partly by individuals in a group constructing their own identities in parallel: through their supposed shared characteristics, again often helped by media orchestration.

This last part of the creation of an imagined community, the creation of a shared identity, is often greatly influenced by a shared history. The role of narrative in constructing group identity and in constructing personal identity is similar. Groups, like individuals, tell narratives about themselves and these narratives reinforce the conceptions they have of their own distinctive characteristics.

"Our" distinct features emerge in the narrative by means of the contrast with the features that emerge as typical of other groups. This need not carry any particular emotional coloring: "We fish; they are farmers". But, of course, usually the narrative is a colored one: "We trusted them, but they let us down"; "They attacked us and we defended ourselves". It is hard for a narrative to create a distinct identity which a group can be proud of without some implied contrast with the defects of others. It is

a difficult conceptual feat to see ourselves as distinctively honest and brave without seeing our neighbors by comparison as treacherous cowards.

But here I want to focus on two more specific aspects of the link between narratives and conflict. One is the role of remembered harm, insult and—especially—humiliation. The other is a third illusion, perhaps worth adding to Amartya Sen's illusions of the simplicity of identity and of destiny. This is the illusion of collective responsibility.

II.3 Remembered Harm, Insult and Humiliation: The Backlash

Remembered harm, betrayal, insult and humiliation often play an immensely powerful role in the narratives that motivate conflict. In the aftermath of Yugoslavia, Serbs, Croats and Bosnians made repeated references to episodes in their group narratives. Each group would remember being betrayed, ruled over, defeated or massacred by one of the other groups. The memories went back centuries. Some Serbs who commanded the war against Croatia, or who ran concentration camps for Croatians, had seen their own parents killed by Croatian fascists, or had been children in Croatian concentration camps. Serbian atrocities against Croatians in the 1990s were partly a backlash against these remembered atrocities.

The backlash against the experience of defeat and occupation is apparent in German nationalism. The influential German nationalist philosopher, Johann Gottlieb Fichte, gave the lectures later published as *Addresses to the German Nation* in French-occupied Berlin in 1807 (Fichte 1807/1968.) Friedrich Schiller noted the response to defeat. His comparison was with walking through young trees that have bendy branches which spring back after being pushed aside. Isaiah Berlin, expounding this idea, calls it "the bent twig" (Berlin 1990), although "backlash" comes closer to suggesting the strength of the response.

The backlash was most striking after the defeat and harsh peace terms of 1918–19, with the forced acceptance of the humiliating "war guilt clauses". It is impossible to read *Mein Kampf* without noticing Hitler's constant boiling anger at all this. His rage was directed against the victorious Allies, but he also believed defeat was the result of treachery: the "stab in the back" for which he made Jews the scapegoat. The major theme of his life's work after 1918 was to undo and avenge these humiliations.

The strength of the backlash may depend on effective propaganda. National revival to expunge the humiliation of 1918 was central to the Nazi message. It was a major theme of the Nuremberg rallies as portrayed by Leni Riefenstahl in *Triumph of the Will*. But successful propaganda needs to connect with feelings and thoughts people already have. This theme clearly found an echo in public opinion.

Part of the backlash against humiliation is the desire to retaliate in kind. After defeating France, Hitler devised a theatrical humiliation. He made the French

representatives sign the armistice at Compiègne where the Germans had signed after their defeat in 1918. The same railway carriage was used, with Hitler in the seat previously occupied by Marshal Foch.

II.4 The Backlash Now

The backlash against perceived humiliation is also central to some of the conflicts of our own time, notably the conflict of which 9/11 is a part, seen on one side as *jihad* and on the other as a war on terror. Neither side has noticed that the other side's actions are part of a backlash.

About a month after 9/11, Osama bin Laden made a statement broadcast by al-Jazeera:

> What America is tasting today is but a fraction of what we have tasted for decades. For over eighty years our *umma* has endured this humiliation and contempt. Its sons have been killed, its blood has been shed, its holy sanctuaries have been violated, all in a manner contrary to that revealed by God, without anyone listening or responding . . . But when after eighty years the sword comes down on America, the hypocrites rise up to lament these killers who have scorned the blood, honor, and holy places of Muslims. (Bin Laden 2005: 104)

Well before 9/11, Bin Laden saw terrorist attacks as a retaliation that would expunge the sense of humiliation. In 1997 he said, "We look upon those heroes, those men who undertook to kill the American occupiers in Riyadh and Khobar. We describe them as heroes and as men. They have wiped disgrace and submissiveness off the forehead of their nation" (2005: 51).

Bin Laden indicates that much of the sense of humiliation comes from parts of the Islamic world being occupied by American troops. As a result, he makes driving them out a central political goal. In 1998, he outlined the aims of the "World Islamic Front": "To kill the Americans and their allies—civilian and military—is an individual duty incumbent on every Muslim in all countries, in order to liberate the al-Aqsa Mosque and the Holy Mosque from their grip, so that their armies leave all the territory of Islam, defeated, broken, and unable to threaten any Muslim" (2005: 61).

The hope that the Americans will leave defeated and broken reflects a familiar part of the backlash: the desire to repay humiliation in kind. For Bin Laden, this was one of the successes of 9/11. Looking back on the attackers, he said, "They rubbed America's nose in the dirt, and wiped its arrogance in the mud. As the twin towers of New York collapsed, something even greater and more enormous collapsed with them: the myth of the great America and the myth of democracy . . . It also became clear to people that America, this unjust power, can be struck down and humiliated" (2005: 194–5).

This same sense of 9/11 being a national humiliation was felt by many Americans, although they did not see it as destroying "the myth of democracy", but as

something to be expunged and avenged. President Bush caught the mood when, speaking to a crowd at the ruined World Trade Center, he said, "I can hear you. The rest of the world hears you. And the people who knocked these buildings down will hear us all soon."

The desire to repay humiliation in kind is visible in the war on terror. The war in Afghanistan was partly an attempt to stop Osama bin Laden and others in al-Qaeda from further terrorism by catching them. But there was also an expressive element to it. Dropping bombs with the initials of the New York Police Department on them must have been seen as part of ensuring that they "will hear us all soon". Saddam Hussein was not involved in 9/11. But when he was captured, unkempt after living in his hole in the ground, to film him while the inside of his mouth was forcibly examined, was a kind of "retaliatory" humiliation.

II.5 The Illusion of Collective Responsibility

It is not just our own nation or group that forms an "imagined community". We also project on to other groups a similar (though usually more negative) imaginary homogeneity. Especially when we are remembering some harm or humiliation inflicted on "us", we are prey to the illusion of collective responsibility. "They" (all of them!) have these bad characteristics, which were expressed in what they did to us, and so they are all responsible. They all deserve what they get as a result.

This illusion is found in the leaders. When Osama bin Laden celebrates 9/11, it is "America" whose nose is rubbed in the dirt. But is it likely that, of the some 3,000 people he killed, all were supporters of American government interventions in the Islamic world? Did the Islamic Americans he killed deserve to have their noses "rubbed in the dirt"? What about the 86 people he killed who came from Japan, Colombia, Jamaica, Mexico and the Philippines? The celebration depends on belief in collective responsibility, which in turn depends on not asking these questions about the actual people who get killed.

And when President Bush targeted Afghanistan and Iraq as part of his "war on terror", he was delivering on his promise that the people who knocked these buildings down "will hear us all soon". But how many of the more than 3,000 Afghan civilians killed does this fit? Were there really more people in the 9/11 conspiracy than were killed by it? And the more than 600,000 people, according to some estimates, killed as a result of the Iraq war obviously were not all involved, and perhaps none of them were. Of course other motives were in play in the Iraq war: possibly to do with oil, possibly to do with gaining political dominance in the Middle East, possibly to do with removing a cruel dictator and installing democracy, and possibly to do with weapons of mass destruction. But one reason for early American support of the war was that it was hitting back at Iraq's supposed implication in al-Qaeda terrorism. Among the false beliefs behind this thinking was the illusion of collective responsibility.

This illusion is found not only in the leaders, but also in those who carry out the supposed acts of retribution. It is found in the videotape made by Mohammad Sidique Khan, one of the terrorist bombers who struck London on 7 July 2005. He was both British and Islamic, but his words made it clear that his chosen identity was Islamic rather than British. "You" are the people of Britain and other Western countries involved in the Iraq war. "We" are Islamic people.

He said,

> our words have no impact upon you, therefore I'm going to talk to you in a language that you understand. Our words are dead until we give them life with our blood . . . Your democratically elected governments continuously perpetrate atrocities against my people all over the world. And your support of them makes you directly responsible, just as I am directly responsible for protecting and avenging my Muslim brothers and sisters. Until we feel security, you will be our targets. And until you stop the bombing, gassing, imprisonment and torture of my people we will not stop this fight. We are at war and I am a soldier. Now you too will taste the reality of this situation. (Khan 2005)

The illusion of collective responsibility is linked to the idea that all members of the other group have shared (usually negative) characteristics. Here "you" (British people) are all impervious to the words of Islamic people and so "blood" is chosen as "a language you understand". "Your" support of governments perpetrating atrocities against Islamic people "makes you directly responsible".

With his bomb at Edgware Road Station, Mohammad Sidique Khan killed (apart from himself) six people. One question he might have asked about his potential victims is whether some of them might have been British *and* Islamic. Another is whether it was likely that all those he killed would be among the minority in Britain who supported the Iraq war. Another is whether the "you" he holds "directly responsible" would include those who took part in London's (largest ever) political demonstration: the one against that war just before it started. The illusion of collective responsibility does not easily survive such questions.

The same illusion helps bring about the kinds of action that fed Mohammad Sidique Khan's resentment. American troops back from Iraq have started to describe atrocities in which some of them participated, together with the psychology behind them. According to one report,

> The mounting frustration of fighting an elusive enemy and the devastating effect of roadside bombs, with their steady toll of American dead and wounded, led many troops to declare open war on all Iraqis. Veterans described reckless firing once they left their compounds. Some shot holes into cans of gasoline being sold along the roadside and then tossed grenades into the pools of gas to set them ablaze. Others opened fire on children.
>
> (Hedges and Al-Arian 2007: 3)

Some of the soldiers were troubled when, back in America, they thought about how they had felt and acted. One said that the general attitude was "A dead Iraqi is just another dead Iraqi" and explained this: "The soldiers honestly thought we were trying to help the people and they were mad because it was almost like a betrayal.

Like here we are trying to help you, here I am, you know, thousands of miles away from home and my family, and I have to be here for a year and work every day on these missions. Well, we're trying to help you and you just turn around and try to kill us." Another said that "the frustration that resulted from our inability to get back at those who were attacking us led to tactics designed simply to punish the local population that was supporting them" (Hedges and Al-Arian 2007: 3, 4).

The illusion of collective responsibility behind waging open war on all Iraqis, and behind wanting to punish the local population, again relies on the collective "you". Again the illusion would not survive such questions as whether the children fired on are really part of the "you" who "just turn around and try to kill us".

The cycle of humiliation and retaliation, together with the linked illusion of collective responsibility, has marked the current conflict. Prisoners of one side may be beheaded, with the atrocity broadcast to the world. Prisoners of the other side may have their religion insulted and may themselves be tortured and sexually humiliated. It is easy to understand the feelings on each side of rage and vengeance. It is less easy to see how to break out of this cycle, created by those on both sides who do not understand how it works.

III. The Enlightenment and the Psychology of Violence

Amartya Sen is right to stress the enormous harm done by the illusions about identity that he has identified. He is also right in exposing these illusions to rational criticism in the hope of weakening people's susceptibility to them. This kind of questioning and criticism has been the task of philosophy at least since Socrates. The line comes down to us through, among many others, the Stoics, Spinoza, Hume, Kant, Mill and Russell. It was particularly at the time of the Enlightenment that the tradition of rational criticism took on a strongly social dimension. Hume, Adam Smith, Diderot, Condorcet, Kant and others, despite their differences on the relative role of reason and the emotions, shared the belief that reason could change the world. There was optimism that challenging the illusions of religion could lead to new and better foundations for morality and politics. Sen's rational arguments against the illusions of identity are squarely in this great tradition.

In some ways, my proposed additions to his account of these illusions and violence present a darker picture. Sen is right that we do not have to accept an identity mainly conceived in terms of religion, nationality or ethnicity. But, if the suggestion here about the role of some of these categories in our psychology is right, even the strongest skeptical arguments may have limited effectiveness. (This does not mean

we should give up making the rational case against the grip of these entrenched categories.) The other suggestion is about the cycle of mutual humiliation and backlash: about how easy it is for groups to get trapped in this cycle and how hard it is to escape from it. This too supports a less optimistic picture.

At this point, I want briefly to add something to a continuing conversation Amartya Sen and I have been having over the last few years. The conversation is about the Enlightenment, and about reason and the emotions. His article "The Reach of Reason", first published in the *New York Review of Books* in 2000 (repr. in Sen 2005), was in part a generous review of my book *Humanity: A Moral History of the Twentieth Century*. At one point, he took me to task for being unfair to the Enlightenment thinkers, including the book in the class of "attacks on ethics based on reason" (Sen 2005: 277). My book laid great emphasis on the importance of sympathy and imagination as barriers against atrocity. Sen justifiably suggested that I should have considered the emphasis given to sympathy and other emotional responses by David Hume and Adam Smith, before criticizing "the thin, mechanical psychology of the Enlightenment". He also criticized my comment that "Stalin and his heirs were in thrall to the Enlightenment".

He then made the point that recently the Enlightenment perspective has come under severe attack—he mentioned John Gray—and suggested that I had added my voice to this reproach. When I read this I was a bit dismayed that I had expressed my intentions so unclearly that even a sympathetic and acute critic saw them in this way.

The point I had hoped to make about Stalin and his heirs, such as Pol Pot, was about overconfidence in their ability to produce blueprints of a totally redesigned society. Like many, I am inspired by Vico's thought that "the social world is certainly the work of men", with its implication that the way a society is set up does not have to be seen as part of the order of nature, but can be changed in ways that may allow people to be less cramped and more fulfilled. Perhaps Vico is only controversially an Enlightenment thinker, but this idea seems to me a core Enlightenment idea, and one that is visible in the Communist project.

My criticism is that the *total* redesign of a society is too grandiose. Such a project is dangerously overconfident and carries with it the risk of things going wrong on a huge scale. And this is especially so when the blueprint is taken as dogma, rather than as something experimental, to be changed or abandoned in the light of feedback that things are not going well. This attack on crude attempts at the rational reconstruction of society was not intended as an attack on reason itself. As I said in my book, "Rationality is saved by its own open-endedness. If a strategy of following accepted rules of rationality is sometimes self-defeating, this is not the end. We revise the rules to take account of this, so producing a higher-order rational strategy" (Glover 1999: 232).

Like Amartya Sen, at a deep level I am on the side of the Enlightenment. The attempt to free morality and politics from religion is one of the great advances

in the history of human consciousness. And the idea that we can create a better and more humane world through rational criticism of enslaving prejudices and illusions is one that Sen and I both support. But the complications I have tried to add here to his picture, and the study of the psychology of atrocities presented in the book he reviewed, are intended to suggest that the obstacles (whether coming from overconfident versions of rationalism or from dark, irrational parts of our nature) may be more daunting than Enlightenment optimism suggests.

The other issue is about whether the psychological views of the Enlightenment thinkers are too simple to give adequate guidance in morals and politics. Adam Smith did indeed write with sensitivity and subtlety about sympathy and imagination. But I do not think he saw the horrors that might come from the darker side of our psychology: from our intermittent love of cruelty, our obedience, our conformity, and our sometimes rigid ideological commitments. Nor did he see the deep roots of our tribal commitments or the psychological importance of defeat and humiliation in generating a self-sustaining cycle of violence.

If the point about humiliation and the cycle of violence is right, it does darken the Enlightenment picture of human nature and of our prospects. But, at the same time, it gives a new argument in support of one of the most visionary political proposals of the Enlightenment: Immanuel Kant's proposal for a world federation of states, which he thought should be given a monopoly of military force (Kant 1970).

The problem of how to police the world and its conflicts is heightened both by modern technology and by our getting closer to being a global village. Because they cause death and suffering on such a scale, the continuing occurrence of war, massacre and genocide has long been a scandal to the human race. But now different groups are so interconnected that the hatreds created by one conflict can create "enemies" all over the globe. An obvious case is the Israeli–Palestinian conflict, whose poison infects so many relations between Western countries and Islamic ones. With the spread of nuclear weapons, and their increasing portability, it may become all too easy for quite small groups to do harm to their supposed enemies in many parts of the world on a much greater scale than that of 9/11.

How should we try to eliminate, or at least contain, violent conflict between groups? Probably the ideal solution would be one in which rational criticism had persuaded us all to give up the illusions on which ethnic, national, political or religious tribalism depend. Those of us who see the illusions for what they are should certainly continue with the long, slow strategy of questioning and argument. Some of the deepest changes in the world are caused by the evolution of beliefs and attitudes, even when this happens with glacial slowness.

But that slowness means that our response to group conflict cannot wait for these long-term psychological changes that may eliminate it. It is so far from being eliminated that we need ways of containing and policing it now. We are often told, and we partly believe, that we live in a global village. How should the global village

respond to the outbreak of violent conflict? Actual villages are of two kinds. In some villages, when violence breaks out people call the police, who have legal authority and enough force to impose the rule of law. In the other kind of village, there is no impartial authority strong enough to prevent violence. The only way to stop it is for the strongest and best-armed neighbors to form a posse and to use their force to keep the peace. Our world hovers between partial attempts to become the law-governed village and what most often is the reality of the lawless one.

Kant's idea is the model of the law-governed global village. He saw that the strong desire of so many groups to have their own nation-state makes a complete world government unrealistic. Instead, he advocated a federation of all the nation-states. The states would give the world federation a monopoly of the use of military force, parallel to the monopoly of force given to the police in a law-governed village, city or state. The United Nations is the embryonic form of such a scheme, but at present it falls far short of it. To turn it into the Kantian world federation, there would need to be a legal constitution spelling out when it should or should not use military force, together with a world court to authorize the use of force in particular cases. And nation-states would have to give up having independent military forces.

In our present world, Kant's idea still seems utopian to most people who hear of it. There is an influential school of political thinkers who argue that, in international affairs, it is impossible or undesirable to have any genuine rule of impartial law. In the view of these self-styled "realists", peace can only be kept by the domination of the strongest power or group. In the global village violence by others is stamped out by the strongest family (currently America), either on its own or by rounding up a posse (a "coalition of the willing").

Among the weaknesses of this approach are issues of authority and impartiality. The world has never asked America, or its allies, or NATO, to be world policeman. Interventions by a self-appointed country or group of countries will be seen as self-interested. Those intervening will also be seen as lawless: not as the police but as an armed gang. (The Iraq war gives some suppport to this.) Even if violence by others is successfully defeated, this is unlikely to be the end of the matter. If the "backlash" psychology outlined here is right, the result is more likely to be vendetta and a cycle of violence. (The claim that the Iraq war has created more terrorists is relevant here.) No doubt police forces are imperfect and what they do is often resented. But, where a police force manages a decent approximation to impartiality, vendettas against them are far less likely than vendettas against rival gangs.

It has been argued here that humiliation and the backlash against it are *central* to violent conflict. This makes it highly likely that any intervention by nations or groups of nations to impose solutions will in turn generate a backlash. Without the legitimacy given by being part of an accepted international authority, those imposing a solution can hardly avoid being seen as victors and occupiers. Lasting peace may not survive the propensity of the defeated and the occupied to hit

back. International authorities, like police forces, have their own problems. But the present state of the world makes vivid the possibly greater problems of a world without any legitimately enforceable rule of law.

Perhaps it is only a short-term and narrow version of realism that prefers the posse to the police. A darker view of human psychology may not show that Hobbes, Machiavelli and Clausewitz were wiser than Kant. The pessimistic view sketched here of the strength of human irrational passions may, surprisingly, add to the rational case for one of the Enlightenment's finest and most optimistic ideas.

IV. Endnote

I first got to know Amartya Sen when we both worked in Oxford and used to talk as we traveled back together on the train to our homes in London. Sometimes he generously gave me a ride back from Paddington Station to my house in his car, driven at hair-raising speed. One day, before he had published *Poverty and Famines*, he opened my eyes to the fact that famines typically happen when there is enough food for everyone to eat, but when many people are denied access to it, often through poverty. This was one of a number of the most stimulating conversations of my life that took place on train journeys with this amiably argumentative Indian.

Of course it is against the spirit of his work simply to classify him as an argumentative Indian. Parts of his identity include being an economist, a philosopher, a parent, a Harvard professor, an egalitarian, a fan of Satyajit Ray, an inspiring teacher, a Nobel Prize winner, a former Master of Trinity College, Cambridge, a member (according to his Hindu grandfather) of the atheistic part of the Hindu spectrum, an unfailingly courteous listener, and a slightly alarming driver.

Despite all this, I will settle here for the "argumentative Indian" phrase I have borrowed from him (though of course he does not use it about himself). The emphasis here is on the adjective rather than the noun. Consider some of the results of the arguing: broadening and humanizing the foundations of economics; persuasively reviving the case against economic inequality; providing a strong alternative to utilitarian thinking in development economics and elsewhere; transforming our view of how to prevent famine; inventing a more flexible version of rights theory; using demographic figures to show in detail how women are failed by many cultures; arguing against deference to the cultural traditions responsible for these failures; producing a powerful case for political freedom as promoting rather than hindering economic development; challenging the illusions about identity that contribute to violent conflict. It is not a bad demonstration of what one argumentative Indian can do.

References

Anderson, Benedict (1983), *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso).

Berlin, Isaiah (1990), "The Bent Twig: On the Rise of Nationalism", in Isaiah Berlin, *The Crooked Timber of Humanity: Chapters in the History of Ideas* (London: John Murray), 238–61.

Bin Laden, Osama (2005), *Messages to the World. The Statements of Osama bin Laden*, ed. Bruce Lawrence, tr. James Howarth (London and New York: Verso.)

Dawkins, Richard (2006), *The God Delusion* (New York: Bantam).

Fichte, Johann Gottlieb (1807/1968), *Addresses to the German Nation*, ed. George A. Kelly (New York: Harper and Row).

Glover, Jonathan (1999), *Humanity: A Moral History of the Twentieth Century* (London: Jonathan Cape).

Hedges, Chris, and Al-Arian, Laila (2007), "The Other War: Iraq Vets Bear Witness", *The Nation*, 30 July.

Kant, Immanuel (1970), *Perpetual Peace*, in *Kant: Political Writings*, ed. Hans Reiss, tr. H. B. Nisbet (Cambridge: Cambridge University Press).

Khan, Mohammad Sidique (2005), "We are at War…I am a Soldier", avaliable at <www.informationclearinghouse.info/article10079.htm>.

Rawls, John (1971), *A Theory of Justice* (Cambridge, Mass.: Belknap Press).

Rushdie, Salman (2002), *Step Across this Line: Collected Non-Fiction 1992–2002* (London: Jonathan Cape).

Sen, Amartya (1977), "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, 6: 317–44.

—— (2002). *Rationality and Freedom* (Cambridge, Mass.: Belknap Press).

—— (2005), *The Argumentative Indian* (London: Allen Lane).

—— (2006), *Identity and Violence* (London: Allen Lane).

Shaw, Paul R., and Wong, Yuwa (1988), *Genetic Seeds of Warfare* (London: Routledge).

Strawson, Galen (2005), "Against Narrativity", in Galen Strawson (ed.), *The Self?* (Oxford: Blackwell), 63–86.

Zimbardo, Philip (2007), *The Lucifer Effect: How Good People Turn Evil* (London: Rider).

CHAPTER 24

FREEDOM AND EQUALITY

FROM IQBAL'S PHILOSOPHY TO SEN'S ETHICAL CONCERNS

AYESHA JALAL

In *Ilm-al-Iqtisad* (The Subject of Economics), his first and little-known book published in 1904, the poet and philosopher Muhammad Iqbal (1877–1938) wondered whether poverty was intrinsic to the natural order of things. "Is it not possible for every human being to be free of the pain of poverty?" he asked. Could the hushed sounds of suffering coming from the streets and alleys be silenced once and for all so that the heart-wrenching scenes of human destitution and degradation were wiped off the face of the earth? The eradication of poverty depended to a large extent on the ethical capabilities of human beings and was, therefore, beyond the scope of economics. A knowledge of economics was nevertheless essential for even the beginning of a solution to the problem of poverty. This was especially true of India, where rampant poverty co-existed with widespread illiteracy. Ignorant about the reasons for their economic ailments, Indians had been unable to initiate basic social welfare schemes. Human history was testimony to the terrible fate that awaited nations oblivious to their cultural and economic conditions (Iqbal 1904/2004: 21–2).

Iqbal's endeavor to learn the rudimentary principles of economics was motivated by pragmatism quite as much as by the desire to discover the root causes of human suffering. After doing a B.A. in English, Arabic and philosophy, he completed his Master's degree in philosophy. He wrote *Ilm al-Iqtisad* while teaching at the University Oriental College in Lahore, where, in addition to translating several English works into Urdu, he taught various subjects, including economics. A compendium of lecture notes with hints of Marxist influence which he later denied for fear of being declared an Islamic heretic, the book is less important for its content than for the insights it provides into Iqbal's early intellectual preoccupations. By the time of his departure for England in 1905, he was a rising star in literary circles, lauded for his evocative Urdu poetry about love for the territorial homeland. But his exploratory foray into the prohibitive world of theoretical economics had gone unnoticed, even though *Ilm al-Iqtisad* appeared in the same year as his hugely popular *Tarana-i-Hindi*, which remains the best-known patriotic narration of the ideal of the Indian nation. While studying at Trinity College, Cambridge, Iqbal is said to have been concerned about becoming too immersed in philosophy. Looking for intellectual breadth to better balance his personality, he spent hours in the University Library reading books on economics (Iqbal 1904/2004: 7). The effort bore fruit not in any further writings on economics but in poems composed over a long and distinguished literary career. In these poems Iqbal blended his understanding of economics with observations on life through the shifting lens of a God-centered humanism that was Muslim and also uniquely his own.

I. Reason, Religion and Ethics

Apart from sharing the same alma mater of Cambridge University, a diehard rationalist wary of religion like Amartya Sen may not appear to have much in common with the poet and philosopher of the East who self-consciously steeped himself in his Muslim identity. But then the badge of greatness worn by two of South Asia's best-known thinkers is an important commonality, as are their broad-ranging and overlapping intellectual interests in philosophy, economics, politics and nationalism. The philosophy and economics that molded Iqbal's thinking in the first decade of the twentieth century were markedly different from what Sen mastered half a century later. Yet there are some broad and interesting parallels in their philosophical conceptions of human freedom and equality that merit identification and elucidation. Instead of being a barrier, their divergent stands on religion can lend a fruitful tension to a consideration of their respective approaches to the problem of poverty and inequality in the modern world.

Iqbal's philosophy of freedom and equality was motivated by many of the same ethical concerns that prompted Amartya Sen to write *Development as Freedom* (Sen 1999). Since everyone wants the capability to live a good life that they can truly value, Sen calls in this book for an expansive view of the idea of development as human freedom rather than simply economic growth. On this view, the quality of life should not be measured by wealth or the availability of constitutional freedoms but by the quality of health, education, life expectancy and civic life. In order for the good life based on genuine freedom to be realized, development policies have to be aimed at enhancing the capabilities of people so that they can influence politicians and bureaucrats. Only when its impact on politics is fully exploited through the empowerment of the people does development contribute to human freedom.

What he saw of life in the advanced capitalist world during his three-year sojourn in Europe led Iqbal to broadly similar conclusions about the imperative of basing his ethics on the foundations of freedom and equality. Yet he launched a far more stringent critique of Western materialism than anything that Sen would embrace in his conception of a good and fulfilling life. Some of this difference must be explained by the context of pre-First World War Europe rather than any intrinsic contradiction between Iqbal's emphasis on religion and Sen's on reason. Alienated by the materialistic excesses of early twentieth-century Europe, Iqbal wrote, in a poem in his anthology *Bang-i-Dara*:

> Men of the West! The country of God is not a shop.
> What you take to be real is a counterfeit coin.
> Your civilization will commit suicide with its own dagger.
> A nest built on a fragile branch cannot last long.
>
> (Iqbal n.d: 169–70)[1]

Written in 1907, the poem is one of the earliest expressions of Iqbal's disquiet with the epidemic of "isms" that, together with the mechanization of life in Europe, constrained his sense of individual freedom. He was repulsed by capitalism's merciless exploitation of labor and thoroughly put off by the signs of excessive rationalism creeping into European thought along with the menacing tide of nationalistic bigotry.

Embarking with aplomb on the pursuit of knowledge, Iqbal, after completing his studies at Cambridge, went on to obtain a doctorate in philosophy from the University of Heidelberg. Given his humble middle-class background, the degree at best ensured him a professorial position with a modest salary. Looking for something more, Iqbal followed the example of several of his countrymen by doing his bar-at-law in London. He also taught at the London School of Commerce and passed an honors examination in economics and political science. Feeling relatively

[1] Unless otherwise stated in the References, translations are my own.

more empowered with his educational investments, Iqbal returned to India in 1908 and started a law practice in Lahore. But his first love was poetry, with philosophy a distant second. Frustrated at having to work for a living, he channeled his creative energies into poetry, fulminating against Western materialism, rationalism and imperialism with the help of Islamic idioms and themes.

In April 1908, Iqbal gave his maiden lecture in English at Lahore, in which he spelled out what he believed were the principal ethical and political ideals of Islam. Speaking as a critical student of comparative religion rather than a teacher or expositor of Islam, he tried approaching the subject from a "thoroughly humane standpoint". He did not "doubt the fact of Divine Revelation as the final basis of Religion" but "prefer[red] to employ expressions of a more scientific content" (Iqbal 1908: 1D).[2] The nub of his argument was that of all the religious systems, Islam took the most positive view of the human capacity to transform the world for the better. This was unlike the life-denying pessimism of Buddhism, which portrays human beings as utterly helpless in the face of universal pain. Christianity too sees nature as evil and considers human beings to be inherently prone to sin. Zoroastrianism views the world as a constant struggle between the forces of good and evil. While taking pain, sin and the struggle against evil seriously, the Qur'an does not teach Muslims to look upon the world as a painful place where only sin and evil can prevail: "Islam believes in the efficacy of well-directed action" and the standpoint of Islam must be described as "melioristic", which is "the ultimate presupposition of all human effort at scientific discovery and social progress" (2E).

For Iqbal, the central proposition regulating the structure of Islam is that there is fear in nature which human beings must overcome. Islam regards human beings as naturally good and peaceful and seeks to free them from this fear by inviting them to put their faith in God: "*The highest stage of man's ethical progress is reached when he becomes absolutely free from fear and grief*" (2E; italics in original). Islam's main ethical ideal, therefore, is to empower man with the will to knowledge so that he can gradually transcend his fear of nature. It is this ceaseless effort towards ethical realization that gives human beings their sense of personality and makes them aware of themselves as a source of power. Whatever "intensifies the sense of individuality in man is good, and that which enfeebles it is bad" (2H). With soaring optimism, Iqbal declared: "Give a man a keen sense of respect for his own personality, and let him move fearless and free in the immensity of God's earth, and he shall respect the personalities of others and become perfectly virtuous" (2H).

While Buddhism and Christianity regarded self-renunciation, poverty and other forms of unworldliness as virtues, Islam saw them as weakening human individuality. Poverty in Islam is a vice—the Qur'an (28.77) exhorts Muslims not to forget

[2] References to this lecture are by paragraph number of the Internet version.

their share in this world (2I). As free, equal and responsible individuals, human beings are the makers of their own destiny and require no mediator between themselves and God. Islam rejects the Christian doctrines of redemption and the infallibility of the Church because these assume the "insufficiency of human personality" and create a dependency that Iqbal saw as "obstructing the ethical progress of man" (2J). The ethical ideal of Islam in a nutshell is "a strong Will in a strong Body" (3E).

Unfortunately Indian Muslims had neither the strength of character nor the physical prowess to attain the ethical ideal laid down for them by Islam. Iqbal ascribed this to the visible absence of "life-force" due to the "decay of the religious spirit", which along with political causes beyond their control had instilled an abject sense of dependence among Muslims in India. Lacking the spirit of initiative, Muslims with their "indifferent commercial morality" failed miserably in business (4B). The goal of educated Muslims was government service, which in a colonized country like India undermined human individuality (4D). As a result, the community faced ruin: the poor had no capital to invest; the middle classes did not start joint economic ventures, due to mutual distrust; and the rich considered commerce and trade beneath their dignity. "Economic dependence," in Iqbal's opinion, was a "prolific mother of all the various forms of evils" (4D).

There was no immediate solution in sight. Iqbal did not have much faith in "education as understood in this country", but it was the only thing Muslims could fall back on (4E). The real question was, what sort of an education? Anticipating Amartya Sen's capabilities argument, Iqbal noted that an education which had "no direct bearing on the particular type of character . . . you want to develop is absolutely worthless" (4F). The education being imparted in India might ensure its recipients a livelihood if they managed to wangle a job in the upper echelons of the colonial bureaucracy. Yet an education restricted to the few was no solution:

> It is the masses who constitute the backbone of a nation. They ought to be better fed, better housed; and properly educated. Life is not bread and butter alone; it is something more. *It is the healthy character that reflects the national ideal in all its aspects.*
>
> (4F; italics in original)

Educated Muslims were alienated from Islam and knew nothing of their own history. Schools and colleges proceeded on the "false assumption that the ideal of education is the training of human intellect rather than human will" (4H). The superficiality of the educational system in British India was also having an adverse affect on Hindus, who were producing "political idealists whose false reading of history" was causing social disruption (4H). For Indians truly to come into their own, they had to establish educational institutions where their social and historical traditions could be revived: "A living nation is alive" only insofar as it "never forgets

its dead" (4G). Iqbal warned Indian Muslims that the world got rid of nations incapable of resolving their difficulties.

II. Islam, Europe and Modernity

Indian Muslims were in a pitiful condition because they had forgotten their dead as well as the ethical ideals of Islam, their distinguishing mark. Membership in the community of Islam was not based on geography but in identity of belief. The only limitation on the liberty granted to the individual Muslim was the interest of the community as the embodiment of Islam. Democracy is the "best form of government for such a community", Iqbal contended, because its primary ideal is "to let a man develop all the possibilities of his nature by allowing him as much freedom as practicable" (6C). Although democracy was "the most important aspect of Islam as a political ideal", Muslims had done nothing for the political improvement of Asia: they had left it to a Western nation to revitalize Asia politically (6C). This was most unfortunate when seen in light of the two main political ideals of Islam: (1) a "horror of personal authority" as an impediment to the freedom of human individuality, and (2) the principle of equality, which, by acting as a "leveling force", had helped make the early Muslims one of the greatest political powers in the world (6E). Instead of freedom, equality and the intrinsic organic unity of Islam, Indian Muslims in their shackles had "out-Hindued the Hindu himself" by practicing a kind of "double caste system" which left them bitterly divided along sectarian lines (6F).

Although Iqbal acknowledged the disparity between Islamic ideals and Muslim practice, his critics have used it to prepare a long charge sheet against him. A common accusation is that far from elaborating its ideals, he distorted Islam by gratuitously borrowing ideas from Western philosophy. For instance, his celebration of individual freedom is often attributed to his exposure to philosophers like Friedrich Nietzsche. Straddling the worlds of Islamic and Western philosophy with consummate skill, Iqbal could detect commonalities where others found incompatibilities, even as he criticized and reformulated their main tenets in accordance with the needs of his own times. It is one thing to suggest that after studying Western philosophy, which he grasped much better than most of his contemporaries in India, Iqbal was led to highlight neglected aspects of the Islamic tradition. But to suggest that he was insufficiently grounded in the Islamic tradition is grossly unfair. Iqbal never ceased bemoaning the slavish imitation of the West by his co-religionists:

> So blazing hot is the wine of modern civilization,
> The Muslim's earthen body has erupted into flames.
> Borrowed light has turned a speck of dust into a fire-fly,
> How marvelous is the trickery of the resplendent sun!
> New ways have changed the temperament of the youth,
> Such beauty, such awakening, such freedom, such fearlessness!
> . . . .
> What a bundle of pleasures the new life has brought:
> Rivalry, marketing conscience, impatience, greed.
> The Muslim assembly is sparkling in the light of the new candle,
> But my ancient wisdom is telling the moths:
> "O moth! You've derived warmth from the candle's light;
> Try burning in your own fire like me if you have any warmth of heart".
>
> (Iqbal n.d.: 272)

Being rooted in his own tradition did not mean closing his mind to new ideas, far less to the cross-fertilization that had taken place historically between Western and Islamic thought. What irked Iqbal was the failure of Western philosophers to acknowledge their intellectual debt to the early Muslims. So if his nostalgia for the past glories of Islam was a bit excessive, it was for reasons other than false pride:

> Why cry over sovereignty, it was a temporary thing;
> There's no escaping the established principles of the world.
> But those pearls of wisdom, those books of our ancestors,
> Seeing them in Europe, rent the heart asunder.
>
> (Iqbal n.d.: 234)

He longed to awaken his fellow Muslims from their slumber so that they could reclaim their heritage, revitalize their present and build a future that was truly worthy of the teachings of Islam. In his own words:

> How can the transient scene of grief frighten me?
> I am confident of the destiny of my *millat* [nation].
> My world is free of the component of despair;
> The news of complete victory is the secret of my zeal.
> Yes, it is true I keep my eye on the times long gone,
> I tell an old story to the assembly's audience.
> Memory of the past is the elixir for my life,
> My past is the interpretation of my future.
> I keep that uplifting period before me,
> I see tomorrow in the mirror of yesterday.
>
> (Iqbal n.d.: 249–50)

Even when absolved of the unpardonable sin of interpreting Islam in light of Western thought, Iqbal has been reprimanded for misinterpreting the Qur'anic

conception of God's omniscience and its implications for individual human freedom (Fazli 2005). This is a far more serious charge since it calls into question the very basis of Iqbal's theory of individual freedom in Islam. The criticism takes its main cues from two debates within the Islamic tradition: (1) between Mutazilite rationalists and Asharite proponents of predestination, over human free choice versus God's omniscience; and (2) over God's immanence versus God's transcendence. At its simplest, the Mutazilites maintained that God had bestowed on human beings a rational free will to decide between good and evil. The Asharites saw this as a denial of God's omniscience and, in the most extreme expressions of their position, argued that there was nothing human beings could do without the foreknowledge and approval of God. This was the kiss of death to any conception of an Islamic ethics, since there can be no moral responsibility in the absence of free choice. The tussle over God's immanence and transcendence revolved around the question of how similar creation is to its creator. Those who upheld divine immanence rarely implied that God is in the material world, as their detractors alleged in labeling them pantheists, but rather maintained that the world itself is inexplicably immersed in God. Iqbal's position on both these questions is of considerable significance in understanding his attitude towards human freedom and equality.

In his mature philosophy, espoused in seven lectures given between 1929 and 1932 and published as *The Reconstruction of Religious Thought in Islam*, Iqbal avoided entangling himself in the sterile debate on whether God's absolute omniscience is incompatible with human freedom. Instead he focused on the issue of divine knowledge as living creative action, rejecting notions of God's omniscience as either fixed or passive. Iqbal dismissed the theological controversy surrounding predestination as "pure speculation with no eye to the spontaneity of life". The idea of history as a predetermined order of events allowed no room for "novelty and initiation"; it saved God's foreknowledge of the future at the expense of his freedom. Having retrieved the idea of divine freedom as a spontaneous and unplanned creative act, Iqbal reconciled it with individual human freedom by arguing that human freedom was conferred by God so that human beings could share life, power, and freedom with God (Iqbal 1930/1996: 74–5). His detractors have alleged that he deliberately fudged the Qur'anic view of divine foreknowledge in order to construct a philosophical argument in favor of freedom and originality in creation. In their estimation, Iqbal's reconstruction of Islamic religious thought failed because in addition to confusing divine foreknowledge with predestination, he committed the cardinal sin of falling into the trap of pantheism by privileging God's immanence over his transcendence (Fazli 2005).

Others have dismissed these objections on the grounds that the core of Iqbal's philosophy is to be found in his poetry and not in the *Reconstruction* lectures. The truth is that his thought crosses both genres and cannot be arbitrarily compartmentalized. A more forceful line of defense has come from the pre-eminent scholar of

Islam, Fazlur Rahman, who called for Iqbal to be rescued from the "posthumous tyranny of interpretation". Describing him as a "great representative thinker of Islam" and not an interpreter, Rahman noted that Iqbal did not identify the central theme in his thought. This has made him acutely vulnerable to misappropriation by advocates of various ideological positions, "ranging from naked Communism to a class conservatism that is an unmistakable voice from the grave". According to Rahman, Iqbal's philosophical system revolves around the notion of reality as a dynamic, creative and ever-expanding spiritual and moral act. This is precisely the idea of God in the Qur'an, a God who bestows upon human beings the "power-in-righteousness" and the "creativity-for-goodness" (Rahman 1963: 439–40, 443).

Without going into the theological and metaphysical nuances of Iqbal's philosophy, it will suffice to say that the binding thread in his poetic and philosophic corpus is the idea of human freedom centered on *khudi*, literally the self or personality, a term which he uses to refer to the self-conscious and dynamic individual. Giving the Qur'anic view of the story of creation in his third *Reconstruction* lecture, he argued that it had nothing to do with the notion of original sin. The main purpose of the Qur'anic story of creation is "to indicate man's rise from a primitive state of instinctive appetite to the conscious possession of a free self, capable of doubt and disobedience". Man's first act of disobedience was also his first act of free choice, which is why God forgave the transgression. It was Adam's eagerness to know what was forbidden, as are the occult sciences in Islam, that resulted in his being placed in an environment which, however painful, was better suited to the development of his intellectual faculties. This was not a punishment so much as an attempt to defeat Satan's object of keeping man "ignorant of the joy of perpetual growth and expansion". Islam places a high value on the individuality and uniqueness of man, who is given the freedom of choice to interpret and master nature, thereby further amplifying his freedom (Iqbal 1930/1996: 78–80).

Iqbal found it surprising that this "unity of human consciousness which constitutes the centre of human personality never really became a point of interest in the history of Muslim thought". He blamed this on the intellectual orientation of the Greeks, Jews, Zoroastrians and Nestorians, who influenced Muslim theological thought with their dualistic conception of the relationship between body and soul. Islamic Sufism, however, understood the intrinsically ethical nature of this unity of inner experience and found its most dramatic expression in the words of great martyr-saint Mansur al-Hallaj: "I am the creative truth." Iqbal conceded that this sort of mystical experience was not communicable or accessible to everyone. This was all the more reason why modern Muslims had to try to comprehend Islam's teachings on the unity of the human personality, its freedom and capacity to act in a creative and purposeful way to alter reality. The task was a daunting one and the "only course open to us is to approach modern knowledge with a respectful but independent attitude", even if it meant arriving at conclusions that were different from those in the West (87–9).

This was not an overture to cultural relativism of a narrow and bigoted sort. Iqbal believed that Islam, with its conception of the unity of creation (*tawhid*), was meant for all of humanity and not simply Muslims. Islam was the "birth of inductive intellect", he claimed, which "alone makes man master of his environment". Prophecy not only reached its perfection in Islam but found cause to pronounce its own abolition. The reason for this was that "life cannot for ever be kept in leading strings" and to achieve "full self-consciousness, man must finally be thrown back on his own resources". Echoing the secularization thesis, according to which man assumes responsibility for his own life, Iqbal declared that the end of priesthood and hereditary kingship in Islam, the Qur'an's repeated mention of reason and experience as well as the emphasis on nature and history as sources of human knowledge, are "all different aspects of the same idea of finality". This did not mean that there was no room for mystical experience, or that emotion must make way for reason. The inner self and the outer world are both considered sources of knowledge in the Qur'an. By encouraging the development of "an independent critical attitude towards mystic experience" Islam proclaims the end of all personal authority claiming a supernatural origin. The desacralization of individual authority opens up fresh vistas of knowledge in the realm of inner experience. It follows the desacralization of nature implicit in the first half of the Islamic confessional—*la ilaha ilallah* (There is no God but God)—which aims at creating "the spirit of a critical observation of man's outer experience by divesting the forces of nature of that Divine character with which earlier cultures had clothed them" (113).

III. Negation and Affirmation

Foreshadowing by several decades Sen's position against uncritical acceptance of beliefs, Iqbal asserted that the principle of doubt was the beginning of all knowledge (114). And the opening word in the Muslim creed, *la*—literally "there is no God"—was a statement of that doubt. Without the power of negation in the *la*, the affirmation of God in *ilaha ilallah* loses its true meaning. The dialectical tension between the negative and the positive constituted the driving force in human history (see Schimmel 1989: 89–91). Iqbal lamented that Muslims could no longer deploy the power of negation to attain positive self-realization. Pious protestations of God's unity without unified inner action simply confused the form of the Islamic creed with its true substance. In criticizing the self-denying otherworldliness that had crept into the Islamic mystical tradition, Iqbal wanted to see self-affirming and independent-minded individuals engaged in concerted collective action to bring about purposeful and positive change in the world.

Discussing the implications of his philosophical insights, Iqbal noted that the foundational principle of Islam was *tawhid*. The Islamic polity was "only a practical means of making this principle a living factor in the intellectual and emotional life of mankind". It demanded loyalty to God—the ultimate and eternal spiritual basis of all life—not to thrones or any other form of human authority. The Qur'an balanced the immutability of the unity of creation with *ijtihad*, literally "independent reasoning", the principle of movement in Islam allowing Muslims to adjust to social change without discarding divine guidance. Requiring self-concentrated endeavor and just the right balance between inductive reasoning and intuitive knowledge, *ijtihad* was the inner mental struggle through which Muslims could retrieve the ethical virtues embodied in the Qur'anic revelation. It offered the only hope to Muslims at a time when subjugation to Western imperialism had stripped them of freedom to question and say "no" (Iqbal 1930/1996: 129–30).

According to Iqbal, the Turks alone among the Muslims had deployed the principle of *ijtihad* in the service of their religious and political thought by broadening the scope of *ijtihad* with the help of modern philosophical ideas. While lauding their efforts, Iqbal thought the Turks had erred in concluding that the idea of the state rather than religion was dominant in Islam. The spiritual and temporal in Islam were not separate spheres of activities. The "nature of an act, however secular in its import, is determined by the attitude of mind with which the agent does it". As "a single unanalysable reality", Islam rejects the bifurcation of matter and spirit, or of mind and soul, as contrary to man's inherent organic unity: "An act is temporal or profane if it is done in a spirit of detachment from the infinite complexity of life behind it; it is spiritual if it is inspired by that complexity" (135–6).

Modern thought had established beyond any shadow of a doubt that the material world was firmly grounded in the spiritual. There was "no such thing as a profane world". The secular was "sacred in the roots of its being". The Prophet of Islam had put it beautifully: "The whole of this earth is a mosque." A state from the Islamic point of view was needed only to "realize the spiritual in human organization". As the main working idea in Islam, *tawhid* translated into "equality, solidarity and freedom". It was incumbent upon the state "to endeavour to transform these ideal principles" into reality by constructing "a definite human organization" based on Islamic precepts. "It was in this alone", Iqbal claimed, "that the State in Islam is a theocracy." Islam was against any form of government "headed by a representative of God on earth who can always screen his despotic will behind his supposed infallibility". This was what had pitted the Christian state against the Church and a solution had to be found by separating the political from the religious. In Islam there was no established church from which the state could separate. Moreover, "Islam was from the very beginning a civil society, having received from the Quran a set of simple legal principles which, like the twelve tables of the Romans, carried . . . great potentialities of expansion and development by interpretation" (136–8).

By abolishing the caliphate in 1924, the Turkish Grand National Assembly had shown how the right to independent reasoning could be exercised in modern times. A republican form of government was wholly consistent with the spirit of Islam and an imperative of the time. Though unsure about the ultimate fate of the nationalist ideal in the Muslim world, Iqbal was of the view that for now "every Muslim nation must sink into her own deeper self, temporarily focus her vision on herself alone, until all are strong and powerful to form a living family of republics". The idealist in him wanted to think that God was "slowly bringing home . . . the truth that Islam is neither Nationalism nor Imperialism but a League of Nations which recognizes artificial boundaries and racial distinctions for facility of reference only, and not for restricting the social horizon of its members". But the pragmatist in him could see that in the absence of "true and living unity", mere "symbolic overlordship" had little meaning. So he proposed vesting the right of *ijtihad* in an elected Muslim assembly, which "in view of the growth of opposing sects" in Islam was the "only possible form Ijma can take in modern times" (138, 140, 152).

The poet in Iqbal never ceased to be inspired by Islamic teachings on universal brotherhood, which he considered a better ideal for humanity than the arbitrary boundaries drawn by the protagonists of territorial nationalism. Yet he never accepted a false dichotomy between love of country and love of humanity, between patriotism and cosmopolitanism. Appropriating poet-philosophers like Iqbal or Rabindranath Tagore for a version of cosmopolitanism based on abstract reason alone, as has been done in Western writing on the subject (e.g. Nussbaum and Cohen 2002), strips their thought of its richness. What humanity needed, Iqbal contended, was a spiritual emancipation of the individual based on universal principles for the spiritual evolution of human society. Europe had become the "greatest hindrance" to "man's ethical advancement" because its idealistic systems were built on pure reason, devoid of the "fire of living conviction which personal revelation alone can bring". For Muslims, on the other hand, the "spiritual basis of life is a matter of conviction for which even the least enlightened man . . . can lay down his life". Given that the basic idea in Islam was that there would be no further revelation, Muslims "ought to be spiritually one of the most emancipated peoples on earth". But emerging from the "spiritual slavery" of pre-Islamic Arabia, the early Muslim community failed to appreciate the true significance of the idea of the finality of prophecy. If contemporary Muslims were to grasp this point and reconstruct their social life according to the ethical principles outlined in the Qur'an, they might yet attain that "spiritual democracy which is the ultimate aim of Islam" (Iqbal 1930/1996: 156–7).

Critics have scoffed at Iqbal's concept of a "spiritual democracy", describing it as "ornamental prose" aimed at awakening the spirit of struggle among Indian Muslims but devoid of any concrete guidelines on how to establish such a form of government (Ahmed 2002: 87–8). A poetic visionary, Iqbal was clearly more interested in changing hearts and minds than in detailing the nuts and bolts of a system

of government where authority was to be exercised in the name of God rather than ordinary mortals. From the perspective of the history of ideas, the concept of a "spiritual democracy" represents the crystallization of Iqbal's views on human freedom and equality. For a fuller exposition of "spiritual democracy", one has to turn to his poetry, which, contrary to what some of his detractors have suggested, is organically linked to his philosophy. This is evident in Iqbal's treatment of two key personalities in contemporary Muslim history—Sayyid Jamaluddin al-Afghani (1839–1897), the Persian propagandist and political activist, and Said Halim Pasha (1863–1921), the Turkish modernist reformer—in both the *Reconstruction* lectures and his poetic *magnum opus*, the *Javid-nama.*

In the lectures, Iqbal singled out Jamaluddin al-Afghani as the one Muslim whose keen insights into the history of Islamic thought and life, "combined with a broad vision engendered by his wide experience of men and manners . . . made him a living link between the past and the future". If only Afghani had been able to channel his "indefatigable but divided energy" into studying Islam, Muslims, "intellectually speaking, would have been on a much more solid ground today". In the same vein, Iqbal hailed Halim Pasha for recognizing that Islam was "a unity of the eternal verities of freedom, equality and solidarity" which refused to be limited to any specific territory (Iqbal 1930/1996: 89, 137). But the Turkish Grand Vizier also realized that the ethical ideals of Islam had been lost sight of on account of local influences:

> The only alternative open to us, then, is to tear off from Islam the hard crust which has immobilized an essentially dynamic outlook on life, and to rediscover the original verities of freedom, equality, and solidarity with a view to rebuild our moral, social, and political ideals out of their original simplicity and universality. (138)

Looking to rediscover Islam's lost dynamism and universal ethical goals, Iqbal took his poetry to new heights in the *Javid-nama*, which is inspired by the Prophet's ascension to the heavens and Dante's *Divine Comedy*. With the master of mysticism Jalaluddin Rumi as his guide, the poet's imaginary spiritual journey through the heavens brings him into contact with several luminaries with whom he discusses the state of the world in general and that of Muslims in particular. During his stop on the planet Mercury, he encounters his two heroes—al-Afghani and Halim Pasha—with whom he has an engaging conversation on European notions of territorial nationalism, communism, capitalism, imperialism, nationalism and democracy. Afghani, the paragon of Islamic universalism, gives Iqbal a succinct assessment of the contradiction between the unity of body and soul and human freedom, and the artificial distinctions of territorial nationalism:

> The Lord of the West, cunning from head to toe,
> taught the people of religion the concept of Country.

> He thinks of the centre, while you are at discord—
> give up this talk of Syria, Palestine, Iraq!
>
> . . . .
>
> What is religion? To rise up from the face of the dust
> so that the pure soul may become aware of itself!
>
> . . . .
>
> Man of reason, the soul is not contained in dimensions;
> the free man is a stranger to every fetter and chain,
> the free man rails against the dark earth
> for it beseems not the falcon to act like a mouse.
> This handful of earth to which you give the name "country",
> this so-called Egypt, and Iran, and Yemen—
> there is a relationship between a country and its people
> in that it is out of its soil that a nation rises;
> but if you look carefully at this relationship
> you will descry a subtlety finer than a hair.
>
> (Iqbal 1932/1966: 55–6)

Along with the denial of human unity and freedom, Europe had fostered the soul-destroying ideologies of communism and capitalism. In Afghani's estimation, the author of *Das Kapital* was a believer at heart but an infidel in mind. There was implicit truth in Marx's philosophy but he had lost sight of the whole picture:

> … Communism has nothing to do save with the body.
> The religion of that prophet who knew not truth
> is founded upon equality of the belly;
>
> . . . .
>
> Capitalism too is a fattening of the body,
>
> . . . .
>
> The soul of both is impatient and intolerant,
> both of them know not God, and deceive mankind.
> One lives by production, the other by taxation
> and man is a glass caught between these two stones.
> The one puts to rout science, religion, art,
> the other robs body of soul, the hand of bread.
>
> (57)

Continuing the attack, Afghani elaborated upon Islam's teachings on equality and respect for human beings, who were asked to honor body and soul so that they can soar to the loftiest heights of creativity and self-fulfillment. Those who believed in God need no station in life. As servants of God, they were free, needed no slaves and were slaves to none. Their kingdom and laws came from God. While the self-seeking mind only thought of its own welfare, God's revelation considered the welfare and profit of all. When someone other than God decided what was right and wrong, the strong man tyrannized the weak:

What results from the laws and constitutions of kings?
Fat lords of the manor, peasants lean as spindles!
Woe to the constitution of the democracy of Europe!

(62)

Afghani condemned Europe for talking up democracy at home but spreading its imperial tentacles to steal the goods of other countries:

God has called the earth simply our "enjoyment",
this valueless "enjoyment" is gratis, gratis.
You landowner, take a wise hint from me:
take from the land your food and grave, but take it not.

(62–3)

The kingdom of the world belonged to those who only knew the phrase "one God" and lived a life of such poverty that they radiated the light of spirituality. In his message to the Russians inviting them to accept the Qur'anic revelation, Afghani applauded them for breaking the back of imperial rule, as had the Muslims. But the Russian people had to turn this "no" into a living affirmation, otherwise their new world order could not rest on a firm foundation:

Find yourself by abandoning Europe!
If you are apprised by the Westerners' cunning
give up the wolf, take on the lion's trade.
What is wolfishness? The search for food and means;
the Lion of the Lord seeks freedom and death.
Without the Koran, the lion is a wolf;
the poverty of the Koran is the root of empire.
The poverty of the Koran is the mingling of meditation and reason—
I have never seen reason perfect without meditation.

(68)

It was not as if Iqbal failed to see the irony of Afghani's stirring account of the excellences of Islam. As the free-thinking living stream, he interrupted the Persian activist at one stage, exclaiming that if what he said was true then either the Muslims were dead or the Qur'an was dead. Said Halim Pasha used the opportunity to get in a word of his own, commenting that God's religion had been rendered more "shameful than unbelief" because the "mullah is a believer trading in unfaith". Oblivious to the message of the Qur'an, which he considered little more than a fable, and with no share in the wisdom of the Prophet, the mullah was the black-hearted, visionless, tasteless, idle gossiper whose hairsplitting arguments had fragmented the community:

Seminary and mullah, before the secrets of the Book,
are as one blind from birth before the light of the sun.
The infidel's religion is the plotting and planning of Holy War;
the mullah's religion is corruption in the Way of God.

(65)

IV. Conclusion: The Individual and the Collective

To conclude that the visible lack of freedom and equality in the Muslim world for the past several decades is not due to Islam but to Muslims having lost their moorings in a West-intoxicated, materialistic, self-serving and Godless world may seem a trifle too convenient. It will certainly fail to sway those who consider religion to be the major stumbling-block in the way of Muslims joining the rest of the world in the forward march towards enlightenment and "moderation". But Iqbal's case for the revitalization of Islam's ethical ideals on freedom and equality does force one to pause and consider whether invoking the old oppositional framework between religion and secularism is really valid in this context. Without denying Iqbal's calculated projection of his Muslim identity, it would be unfortunate if his use of Islamic terminology were to dissuade us from understanding the logic he espouses with regard to issues of human freedom and equality.

Even as the religious paradigm separates Iqbal from Sen's secularism, there is evidence to suggest that they occupy considerable common ground on issues of freedom and equality. The primary refrain in Iqbal's philosophy and poetry is that of the dynamic, self-affirming individual working in conjunction with the collective for a richer, better balanced and more meaningful human existence. Like Sen, Iqbal advocated a needs-based approach to education, warning against the dangers of treating human beings as machines. Iqbal shared the Nobel laureate's acute concern for poverty and the future of the dispossessed as a living issue and not merely a philosophical one. Those who may be inclined to dismiss much of what Iqbal had to say about freedom and equality as the idealistic musings of a man who was a poet and a philosopher need to be reminded that Iqbal also fancied himself a hands-on politician, eager and willing to tackle poverty and inequality through concrete political action.

It was with this intention that Iqbal wrote to Mohammed Ali Jinnah on 28 May 1937. "The problem of bread is becoming more and more acute", he told the seasoned constitutional lawyer, who had returned from self-imposed exile in London to again take up the leadership of the All-India Muslim League. The Muslim was

beginning to feel that he had been "going down and down during the last 200 years". The common view was that this "poverty is due to Hindu money-lending or capitalism", since the "perception that equality [is?] due to foreign rule has not yet fully come to him". But it was "bound to come" and when it did, the "atheistic socialism of Jawahar Lal [Nehru] [wa]s not likely to receive much response from the Muslims". The burning question of the day, as Iqbal saw it, was how "to solve the problem of Muslim poverty". He was certain that "the whole future of the League depends on the League's activity to solve this question". If the League could "give no such promises", Iqbal was quite certain that "the Muslim masses will remain indifferent to it as before". He thought it fortunate that there was "a solution in the enforcement of the Laws of Islam and its further development in the light of modern ideas". His close, sustained study of Islamic law had led him to believe that "if this system of Law is properly understood and applied, at last the right to subsistence is secured to everybody". But the implementation of Islamic sharia was "impossible in this country without a free Muslim state or states". It had been his "honest conviction for many years" and he still believed that this was "the only way to solve the problem of bread for Muslims as well as to secure a peaceful India" (Allana 1967: 129–30).

In the event Iqbal was more successful in inspiring the movement for a free Muslim state of Pakistan than in solving the problem of poverty and inequality for the majority of the subcontinent's Muslims. This is why Amartya Sen, whose family came from Dhaka in what is now Bangladesh, was perhaps destined to meet the Pakistani economist, Mahbub-ul-Haq, in Cambridge, forge a lifelong friendship and lay the basis of what has come to be variously known as the entitlements, capabilities and human development approach. It is of course another matter whether the hopes and dreams of these men, who pioneered the idea of quality over quantity as the ultimate goal of human development as freedom, remain to be fully realized. That idea, at least, is one undeniable thing that Amartya Sen has in common with the poetic visionary of Pakistan.

References

Ahmed, Ishtiaq (2002), "Looking Backwards into the Future: A Critique of Islamic Modernism", *Journal of Futures Studies*, 7(2): 75–95.

Allana, G. (ed.) (1967), *Pakistan Movement Historical Documents* (Karachi: Department of International Relations, University of Karachi).

Fazli, Abdul Hafeez (2005), "Iqbal's View of Omniscience and Human Freedom", *The Muslim World*, 95(1): 125–45.

Iqbal, Muhammad (1904/2004), *Ilm al-Iqtisad* (repr. Lahore: Sang-e-Meel).

—— (1908), *Islam as an Ethical and a Political Ideal: Iqbal's Maiden English Lecture*, ed. S. Y. Hashimy (Lahore: Orientalia, 1955), 53–101; available at <www.columbia.edu/itc/mealac/pritchett/00islamlinks/txt_iqbal_1908.html>.

—— (1930/1996), *The Reconstruction of Religious Thought in Islam* (repr. Lahore: Sang-e-Meel).

—— (1932/1966), *Javid-nama*, tr. Arthur J. Arberry (London: Allen and Unwin).

—— (n.d.), "Bang-i-Dara", in *Kulliyat-i-Iqbal* (Lahore: Sheikh Muhammad Bashir and Sons).

Nussbaum, Martha C., and Cohen, Joshua (eds) (2002), *For Love of Country* (Boston: Beacon Press).

Rahman, Fazlur (1963), "Iqbal's Idea of the Muslim", *Islamic Studies*, 2(4): 439–45.

Schimmel, Annemarie (1989), *Gabriel's Wing: A Study into the Religious Ideas of Sir Muhammad Iqbal*, 2nd edn (repr. Lahore: Iqbal Academy).

Sen, A. K. (1999), *Development as Human Freedom* (New York: Knopf).

CHAPTER 25

PROTECTIVE SECURITY OR PROTECTION RACKETS? WAR AND SOVEREIGNTY

MARY KALDOR

According to Amartya Sen, protective security is one of the five distinct types of instrumental freedoms that help to advance a person's capabilities.[1] He uses the term to refer to protection against what he calls the "downside risks"—extreme contingencies like famines or financial crises that threaten survival. This notion of security has been applied in the literature about human security, with which Sen is closely associated. The Canadian definition of human security is often conflated with the "responsibility to protect", the doctrine that the international community has a responsibility to override sovereignty in order to protect people in cases of genocide, ethnic cleansing or other large-scale violations of human rights. The broader and more expansive definition contained in *Human Security Now* (Commission on Human Security 2003), by the commission chaired by Amartya Sen and Sadako Ogata, refers to the "protection and empowerment" of individuals and

[1] The other four are: political freedoms, economic facilities, social opportunities and transparency guarantees. See Sen (1999).

communities and it applies not just to the threat of violence but to a range of other economic, environmental or cultural threats as well.

Yet, as Charles Tilly has pointed out, the term "protection" has two quite different meanings:

> One is comforting, the other ominous. With one tone, protection calls up images of the shelter against danger provided by a powerful friend, a large insurance policy, or a sturdy roof. With the other, it evokes the racket in which a local strongman forces merchants to pay tribute to avoid damage, damage the strongman himself threatens to deliver.... Which image the word "protection" brings to mind depends mainly on our assessment of the reality and externality of the threat. Someone who produces both the danger and, at a price, the shield against it is a racketeer. Someone who provides a needed shield but has little control over the danger's appearance qualifies as a legitimate protector. (Tilly 1982: p.1.)

Tilly suggests that the distinction between the state as legitimate protector and the criminal organization as racketeer was built up historically over a long period. In the sixteenth and seventeenth centuries, kings often employed outlaws and privateers and encouraged their soldiers to loot and pillage. Gradually, however, professional forces were established, bandits and pirates were marginalized and even eliminated within the territory of the state, increased taxation and borrowing was used to pay for public services such as the police or education and health, and the "master narrative" of the state, as Ulrich Beck puts it, was constructed around its role in protecting people against risk—the dangers posed by nature and personal risks of ill health and unemployment, as well as threats posed by foreign enemies (Beck 1992). Tilly (1982) suggests that three factors were crucial in building the idea of the state as legitimate protector—economies of scale in the production and use of force which allowed the state to become increasingly effective, the growing assent of the subject population, and international collaboration with other states.

Beck and others suggest that as a consequence of globalization, the state is failing to provide protection. In a world of global risk (climate change, the spread of disease, the uneven effects of the global market, terrorism or transnational crime) the state no longer has the capacity to protect its citizens. I do not disagree with this suggestion but, in this essay, I want to focus on protection from violence. It is generally assumed that all forms of violence within the boundaries of the state are criminal and that the state guarantees internal security through the rule of law and policing. Externally, the state protects its citizens against foreign enemies by using the military. I want to suggest that the distinction between legitimate and illegitimate protection is breaking down because the traditional external role is no longer effective and the distinction between inside and outside is becoming blurred. If states are to provide protective security, this is only possible, nowadays, within a multilateral framework, in which protective security is something that every human being can expect and not just the citizens of some countries. In making

this argument, I will start by describing how contemporary war can be perceived as a protection racket and then consider the ways in which the distinction between legitimate and illegitimate protection is breaking down. In the final section, I will discuss the prospects for reconstructing legitimacy around human or protective security.

I. Contemporary War as a Protection Racket

We tend to think of war on the model of nineteenth- and twentieth-century wars, as wars with two sides, usually states, in which the aim is victory by one side over the other. Such wars are short, immensely destructive and decisive. However, the experience of the Cold War, the War on Terror and the "new wars" taking place in the Middle East, Africa and Central Asia suggests that war has become something different. David Keen refers to war as a "system" (Keen 2007). One could equally well talk about a racket, in which the "sides" are engaged in a sort of collusion to sustain a specific type of political economy. The idea of "protection" is at the core of such a system. I do not want to suggest that war is a deliberate or conscious racket (although conspiracies are often more real than we think). Rather, the system is shaped by a specific context, way of thinking, and set of vested interests.

I have argued elsewhere that the Cold War is best understood as an "imaginary war", which sustained and extended the institutions constructed during the Second World War (Kaldor 1991). The Second World War represented the apex of Tilly's process of state-building during war. In the titanic struggle with the Axis powers, a new bargain was hammered out between the state and its citizens. In exchange for risking their lives and, in the West, for a huge increase in taxation and borrowing, citizens in both West and East gained employment, higher wages (though still higher in the West) and welfare services. But the cost of the war is unbearable even to contemplate. No one knows the true extent of the casualties. Estimates vary between 50 and 70 million dead. Six million Jews, Sinti and Roma, disabled people, homosexuals, and political opponents of the Nazi regime were killed in the Holocaust. Twenty million Soviet citizens died defending their country. Hundreds of thousands died in bombing raids on Dresden, Tokyo and Hiroshima.

The Cold War can be viewed as a way of reproducing the gains of the Second World War without the costs. Over and over again, in exercises on the North German plains, in strategic plans, in spy stories, in hostile rhetoric and in the action–reaction dynamics of the arms race, East and West enacted an imaginary

war, with many features similar to those of the Second World War. It can be argued that it was a mutually self-sustaining system, in which planning was based on worst-case scenarios so that behavior by one side substantiated the threat hypothesis of the other and vice versa. Both sides had an interest in preserving the idea of war. In the West, it was an idea that reminded the world of the American contribution to ending the war and provided "reassurance", to use Michael Howard's phrase, that Western Europe was protected from the Soviet threat (Howard 1982: 3). It justified the continued flow of dollars to Europe, either in the form of aid or of troop presence, despite Congress's objections to the left-wing governments elected all over Europe after the war. And within the US, it provided a way for the Republicans to agree to maintain big government. In the East, the idea of war helped to sustain the notion of socialism as a continual struggle and it provided the rationale for Soviet domination over Eastern Europe, even though bloody methods had to be used in 1953 (GDR), 1956 (Hungary) and 1968 (Czechoslovakia). On both sides, it employed literally millions of men and women in laboratories, factories, and the armed forces and in a myriad of associated services.

Of course, there were real wars outside Europe. Indeed, roughly 5 million people died every decade in wars in the second half of the twentieth century. Whatever the underlying causes, they were viewed as proxy wars, as part of the overall global confrontation between East and West. They helped to substantiate a conceptual world order based on the binary opposition between East and West.

In other words, huge military budgets and a technological arms race offered people protection against an enemy, real or imagined, on both sides and thus sustained a system of world order in which military and industrial institutions were deeply embedded. The existence of each side and each action and reaction can be said to have constructed a threat against which people felt protected.

The War on Terror reproduces the logic of the Cold War. The use of military force to deal with a group of criminal extremists is, on the face of it, irrational and counter-productive. As David Keen (2007: 1) puts it: "Trying to apply the old militaristic model to the problem of terrorism is like trying to destroy a liquid with a sledgehammer or a virus with a bullet." The use of military force in Afghanistan and Iraq has not reduced the terrorist threat. On the contrary, terrorist incidents have increased dramatically, especially in the Middle East and Europe, and the evidence suggests that al-Qaeda cells are multiplying. So why did the Bush Administration launch the War on Terror? Part of the answer is conditioning. Members of the Bush Administration had grown up during the Cold War—their ideas and worldviews were shaped by that experience. Many of them had served in the Reagan Administration. Military responses were what they knew about—their knee-jerk reaction. Moreover, unlike Russia, the United States never had its *perestroika*. Behind the Bush Administration was a complex set of corporate and institutional interests, all of which required military spending for sustenance. And part of the answer is psychological. In an era of budget deficits, inequality, high levels of crime, and

everyday insecurity, the memory of World War II and the Cold War provided for many ordinary Americans a sense of comfort, of being protected.[2]

Precisely because the use of military force seems actually to stimulate terrorism, the War on Terror acquires a permanent justification, much like the Cold War. Every terrorist incident can be used to justify further military action and further defense spending, which, in turn, will provoke further incidents. The palpable relief of the Bush Administration after the uncertainty of the post-Cold War decade is evident in many of their statements. Rumsfeld and Bush talk about a "long war", lasting perhaps 50 years. And according to Cheney,

> When America's great enemy suddenly disappeared, many wondered what new direction our foreign policy would take. We spoke, as always, of long term problems and regional crises throughout the world but there was no single immediate global threat that any roomful of experts could agree upon. All of that changed five months ago [on 9/11]. The threat is known and our role is clear now. (quoted in Keen 2007: 68)

As long as the War on Terror continues to provide psychological reassurance, the racket can continue. The racket can be challenged only to the extent that ordinary Americans begin to question the narrative of protection.

This kind of war system also characterizes the genuine new wars taking place in other parts of the world. Many of these wars are about ethnic or religious identity and commentators often talk about "ancient hatreds". But it can be argued that the idea of "ancient hatreds" is actually constructed through violence. It helps, of course, if there are historic injustices to be cited. But anyone who has witnessed the evolution of today's identity conflicts will have experienced a similar phenomenon—the surprise expressed by the participants in these wars, whether Sunnis and Shi'a in Baghdad, Catholics and Protestants in Belfast, Muslims and Serbs in Sarajevo—that they were friends and able to live together as cosmopolitan citizens before the war began. The experience of violence creates fear and trauma and the need to seek protection from those who are actually creating the enmity through sectarian violence. The narrative of "ancient hatreds" only compounds the fears. As a Bosnian colleague of mine from Banja Luka put it: "This war had to be so bloody because the ties between us were so strong." It was al-Qaeda's attacks on Shi'ite neighborhoods in Baghdad that mobilized the Shi'ite death squads. In other words, through violence, political adventurers are able to mobilize round extreme and exclusivist ideas as a way of gaining power, especially where state institutions are weak.

These wars have an economic dimension as well. During the Cold War, the different sides in the conflict were financed through superpower patrons. In the latest conflicts, the warring parties have to seek alternative sources of finance—looting, pillaging, booty, "taxation" of humanitarian aid, kidnapping and hostage-taking, remittances from the diaspora, or various kinds of smuggling (drugs, cigarettes,

[2] Keen (2007) details many of these psychological functions that the War on Terror fulfills—a sort of magic or witchcraft that allows people to displace their fears and humiliation.

people, oil, diamonds). Indeed, a kind of predatory political economy is established in which it is no longer clear whether the war aims are profit or power. All of these sources of finance depend on violence. The wars provide a framework, a racket, for a kind of primitive accumulation.

Some would argue that war has always been a racket, that what is happening now is no different from earlier wars. Both Rousseau and Kant observed the way in which monarchs emphasized foreign threats in order to consolidate power at home. "War", wrote Rousseau, "furnishes a pretext for exactions of money, and another, no less plausible, for keeping large armies constantly on foot, to hold their people in awe" (quoted in Hoffmann and Fidler 1991: 91). But during these wars, states did construct the institutions of protective security, such as emergency services like the police and the fire brigade. The wars did provide a framework for establishing domestic security. The problem now is that many people experience a renewed insecurity arising from terrorism as well as other global risks like climate change and poverty. There is a gap between what is offered in terms of protection—the war rackets of today—and everyday experience of insecurity. How did this come about?

II. The Breakdown of the Distinction between Legitimate and Illegitimate Protection

By legitimacy, I mean both legality and public consent. Or to put it another way, an institution is legal if it can be justified in a court of law, and legitimate if it can be justified to public opinion. In a law-governed society, legality is a necessary condition for legitimacy even though there are sometimes moments—and these are dangerous moments—when the two do not coincide. Legitimacy could be said to be the outcome of a social contract implicit in the political process.

There are three main reasons for the breakdown of the distinction between legitimate and illegitimate violence: the growing ineffectiveness of military force; the illegitimacy of attacks against civilians; and the re-privatization of organized force, which also means the blurring of the distinction between political violence and organized crime. I will deal with each of these reasons in turn.

II.1 The Growing Ineffectiveness of Military Force

Charles Tilly argues that a key factor in establishing the legitimacy of states was economies of scale in military capabilities that allowed states to build up a

monopoly position within the territory they controlled. The professionalization of troops, the application of technology to weaponry, the improvements in logistics all required substantial funding. Additional resources resulted in disproportionate increases in the effectiveness of military force. As the state's military capabilities became more and more effective, it became increasingly difficult for any organized group apart from other states to challenge them.

It can be argued, however, that during the twentieth century, military technology reached a point of diminishing returns. That is to say, every additional dollar or ruble devoted to increasing military effectiveness led to ever smaller gains in effectiveness. The First World War had already revealed the immense destructiveness of modern military technology and the huge difficulty of securing a decisive victory over a similarly armed opponent. The stalemate of the First World War was eventually overcome by the invention of the tank. The mass production of tanks and aircraft in the Second World War was critical in ensuring the dominance of the United States and the Soviet Union. During the Cold War, the arms race took the form of technological competition to improve the speed, destructiveness, accuracy and protection of what were called weapons systems (tanks, aircraft, warships and rockets). There was, however, a steady increase in the cost of achieving extra increments of speed, destructiveness, accuracy and protection; moreover, any improvements achieved were often associated with costs such as increased complexity, increased logistical requirements, and greater difficulty in operating equipment. The consequence has been a reduction in the scale of the military effort and a reluctance to risk battles in which expensive equipment and expensively trained men and women might be lost.

Meanwhile, small arms—rifles, machine guns, hand grenades, hand-held rockets—have become cheaper, lighter, more easily available and easier to use (Small Arms Survey 2007). Together with the spread of commercially available communications (mobile phones, Internet, air travel), this has meant that the entry costs for non-state armed groups have become much lower. 9/11 is a case in point: a small group of men were able to inflict damage equivalent to a small nuclear weapon by hijacking civilian airliners.

The implication of these developments is two-fold. First, states are much less willing to go to war with other states. Battles have become too destructive and too indeterminate to be fought. The Iran–Iraq war was the exception that proved the rule. The war was a stalemate much like the First World War, in which a million young men were killed after seven long years in which the two armies faced each other across trenches, occasionally taking or taking back very small chunks of territory.

Secondly, non-state armed groups can inflict considerable damage on states. They cannot, of course, defeat the armed forces of states militarily. But they cannot be defeated by states, either. And as all the counter-insurgency manuals point out, as long as they do not lose they are winning. Conventional military force is immensely

destructive. Grozny has been reduced to rubble; yet violence continues. The damage done by Israel during the war with Lebanon in 2006 was extensive—all the bridges on the roads south of Beirut were destroyed and there was widespread destruction of Shi'ite suburbs and villages. But Hezbollah emerged as a victor.

In other words, military force is becoming less effective as a hard instrument of power; it cannot compel others to behave in certain ways. Paradoxically, the power derived from military force may be soft rather than hard, to use the terminology of Joseph Nye. In other words, it rests on appearance, on ritual rather than on physical coercion. It reminds us of past victories and it threatens us with destruction. And as long as the ritual is reassuring, military force will continue to serve some function in the construction of legitimacy. But the more the ritual is challenged, as in Iraq or Afghanistan, the more legitimacy is eroded.

II.2 Civilians as Victims

To say that military force is less effective as a form of hard power does not mean that it is any less destructive. On the contrary, the key characteristic of contemporary wars is that the main victims are civilians (Kaldor 1999). The statistics are notoriously poor but one number often quoted suggests that, at the turn of the twentieth century, 90% of the casualties in war were military and 10% civilian, while at the turn of the twenty-first century that ratio has been precisely reversed. Already in the two world wars, the mobilization of whole societies meant that the distinction between the military and civilians was beginning to break down. In World War I, three quarters of the casualties were still military if the figures for the Ottoman Empire are excluded. In the Ottoman Empire, some 4 million civilians died during World War I, both because of disease and the deliberate genocide of Armenians. In World War II, civilian casualties were much higher, both because of aerial bombing and because of the Holocaust. During that war, some 65% of casualties were civilians.

Today, the share is much higher. In the Iraq War, there have been some 655,000 deaths as a result of the Coalition invasion in 2003 and the subsequent violence up to July 2006. Of these, some 600,000 were the direct result of violence and some 25% were due to Coalition attacks.[3] Most of these Iraqi deaths were civilian. By contrast, some 3,316 American soldiers and 144 British soldiers had been killed up to April 2007 (Iraq Coalition Casualty Count 2007a). In Darfur at least 200,000 civilians have been killed since 2003 through direct violence (UN News Centre 2007). In Afghanistan the number of civilian casualties is not counted. An estimate by the Project on Defence Alternatives suggests that some 3,000 civilians died as a result of air strikes or lack of access to humanitarian assistance during the 2001 invasion,

[3] This estimate was based on a careful analysis of clusters of sample households and in the majority of cases, actual death certificates were produced. See Burnham, Lafta *et al.* (2006).

but many more have died since then. At the time of writing (October 2007) some 707 Coalition soldiers have died since the start of the US-led Operation Enduring Freedom in 2001 (Iraq Coalition Casualty Count 2007b).

One of the most striking characteristics of contemporary political violence is the huge rise in population displacement. According to the United Nations High Commission on Refugees (UNHCR), the global refugee population has increased from 2.4 million in 1975 to around 10 million today; in addition, the UNHCR estimates that there are some 8 million internally displaced persons (IDPs). Other estimates are even higher; the American Refugee Committee claims that there are over 32 million refugees and displaced persons. The number of refugees and internally displaced persons per conflict has been estimated to have increased more than three-fold from 1969 to 2004—from 327,000 to 1,093,300 (UNDP 2006: 12).

In the "new wars" in the Balkans, Africa and the Middle East, the killing of civilians is deliberate. It is a central component of military strategies that have evolved over the past few decades as a way of getting around concentrations of advanced military technology and the diminishing utility of conventional military force. Instead of trying to control territory militarily, by defeating an opposing army, paramilitary groups or a combination of regular and irregular forces try to control territory politically by using violence to control the population. Whereas revolutionaries like Mao Tse-tung or Che Guevara aimed to control territory by winning the support of the local population—by appealing to "hearts and minds"—today the aim is to sow "fear and hatred" by expelling or killing those of a different ethnicity or religion, or even just dissidents and intellectuals. This, for example, is how Serb forces controlled territory in Bosnia. Regular Serb forces would shell a village to create an atmosphere of panic. Then paramilitary groups with names like "Tigers" or "Chetniks" or "Scorpions" would enter the village. Men would be separated from women and killed, especially if they were rich or intellectual, or imprisoned in detention camps. Women would be raped and/or expelled—the argument was that Muslim women would feel so ashamed of being raped that they would never want to return to their homes. Symbols of culture and historic buildings would be destroyed.

Similar tactics are used by the Janjaweed in Darfur and in Chad. Reports speak of systematic and often race-related rape raids, with some women having been branded on the hand in order to stigmatize them permanently. Abductions of girls and boys, random killings of civilians, and the destruction of private property as well as water sources and other essential civilian properties are common (Human Rights Watch 2004). In places like West Africa and Uganda, hideous atrocities have been carried out—like beheadings, throat cutting, and cutting off limbs, ears, lips and noses. In Iraq, Sunni suicide bombers attack crowded Shi'ite areas and Shi'ite militia organize death squads. And a 2007 report by Amnesty International documents a dramatic increase in deliberate attacks on civilians by Afghan insurgent groups, primarily Taliban and Hezb-e Islami forces (Amnesty International 2007).

Suicide bombing carried out by terrorists, as in New York, Madrid and London, or Palestine, Iraq and Afghanistan, is another tactic for avoiding direct confrontations with superior military forces. Suicide bombing is sometimes intended to create fear and panic among the civilian population. It is also sometimes used as a cost-effective way of attacking heavily protected strategic targets (the Twin Towers of the World Trade Center in Manhattan, American forces in Iraq and Afghanistan, or a political figure, like the lake Rajiv Gandhi). But because civilians are unprotected they are usually the main victims.

Counter-terrorism is supposed to counter those tactics. But counter-terrorist tactics actually have the same effects. Because combatants cannot be separated from non-combatants, supposed attacks on insurgents by the US in Iraq, Russia in Chechnya, and Israel in Lebanon and Palestine mainly kill or displace civilians. This is particularly true of air strikes, which are a way of minimizing military casualties while trying to attack strategic targets or combatants. Even though Western governments claim to minimize so-called collateral damage, air strikes are necessarily imprecise, despite advances in technology, and in addition, are terrifying for those on the ground. Indeed, this is the point of tactics like "shock and awe". The use of cluster munitions—bombs that explode into tiny fragments—by the Israelis in Southern Lebanon or Gaza, for example, leaves large amounts of territory unusable and results in casualties, especially children, long after the bombs have been dropped. Where ground forces are used, killing is often arbitrary—some five to ten Iraqi civilians have been killed on average each week by Americans at checkpoints. Thousands of young men are detained and interrogated: Israel currently holds some 10,000 Palestinian prisoners; the US holds around 17,000 Iraqi prisoners and this number is growing (Centre for Research on Globalization 2007). Walls are built to divide people, which often separate families or landholdings or prevent access to work, like the Israeli wall in Palestine, or the new walls being built by the Americans in Baghdad. Casualty levels from counter-terrorism are actually much higher than from terrorism itself. In the 2006 war in Lebanon, for example, the Israelis killed some 1,200 Lebanese civilians; estimates of combatants killed vary from 500 (Israeli defence forces estimate) to 74 Hezbollah fighters (Hezbollah estimates) and 46 Lebanese soldiers (Lebanese figures). Some 44 Israeli civilians died, along with some 119 members of the Israeli Defense Forces.

Wars can be described as legitimate killing. The legitimacy derives not just from the authority of the state but also from rules that have developed over centuries about the conduct of war. These rules were codified in the late nineteenth and twentieth centuries into what is known as international humanitarian law or the "laws of war". These laws cover such issues as the treatment of prisoners, the sick and wounded, and non-combatants, as well as the concept of "military necessity" and the definition of weapons and tactics that do not conform to this concept.

In the period after World War II, war itself was prohibited except in self-defense, thus narrowing further the definition of legitimate killing. In addition, human

rights norms, which apply to the way states treat their own citizens, were codified in the Universal Declaration of Human Rights, the International Covenants on Human Rights, the Genocide Convention, and the Convention on Torture, as well as the European Convention on Human Rights and other conventions dealing with discrimination against women, the rights of the child, and the rights of the disabled.

All killing of civilians is illegal in the international framework of human rights. The killing or detention of civilians can sometimes be justified if it is judged within the framework of international humanitarian law to be "militarily necessary" and proportionate. The anodyne term "collateral damage" is used to justify the killing of civilians, for example, in counter-terror tactics, when the civilians are killed or displaced as a consequence of attacks on what are thought to be combatants. But legality is not always the same as legitimacy, in the sense that those who inflict violence are not necessarily able to justify what they do in a court of law, but rather in terms of public opinion. The "collateral damage" justification may be convincing for Western publics but these kinds of deaths in Iraq, Afghanistan, Palestine and Serbia are viewed by the victims and their communities as violations of human rights or as terror. By the same token, suicide bombings or attacks by insurgents are usually justified in the name of war—they are the only way to attack a stronger enemy. But their victims and their families consider such attacks as terrorism.

Protection rackets thus depend on the extent to which they can be presented in the framework of war and the extent to which war itself is considered legitimate. The more that protection rackets are viewed within the framework of human rights, the more difficult it is to sustain their legitimacy. Precisely because communities nowadays cross borders and because we can witness what happens in faraway places, the distinction between war and terrorism, or between legitimate and illegitimate violence, is being increasingly eroded.

II.3 The Privatization of Organized Force

In the last two decades, private security companies have been increasingly used for policing and protection not only at home but also in overstretched peacekeeping missions and in other military interventions. Coalition forces are heavily dependent on private security companies in both Iraq and Afghanistan. These companies are not subject to the same degree of regulation as military forces and, on occasion, this has mired peacekeeping or intervention forces in difficult situations. The recent case of the shooting of Iraqi civilians by Blackwater is a good example.

"New wars" are fought by networks of state and non-state actors. The latter can include warlords, paramilitary groups, mercenaries, and criminal groups. Both state and non-state actors increasingly finance themselves not through taxation but through various types of illicit activities—the origins of which can often be traced back to authoritarian regimes.

Within the framework of authoritarian regimes, all kinds of independent economic activity are criminalized. There is a tendency to assume that the opening up of such regimes will lead to the establishment of legitimate markets. In fact, the demise of Soviet Communism and South Africa's apartheid system was associated with the explosion of organized crime, which had its roots in the illicit interstices of authoritarianism. In South Africa, the Goldstone Commission and the Truth and Reconciliation Commission revealed the extraordinary range of criminal activities in which security forces were engaged: drug, ivory, precious gem and mineral running; bank robbery, embezzlement and money laundering; falsification of records, illegal sales of licenses, and fabrication of evidence. In addition, of course, they were involved in illegal political violence—human rights abuses, torture and maiming, and political killings as well as fomenting violence and racism. As any visitor to Russia in the last days of Communism will testify, smuggling, black market exchanges, trafficking and other activities were rampant.

What Carolyn Nordstrom (2004) calls "shadow networks" are the underside of the growth of the global market. The "new wars" have provided a fertile zone for these transnational networks of state and non-state actors, who are engaged in a range of illicit activites related to violence. These include: looting, pillaging and hostage-taking; establishing checkpoints so as to control the flow of necessities and charge "customs duties"; smuggling in valuable commodities like oil or diamonds or illicit commodities like drugs, cigarettes, alcohol and people; and "taxation" of humanitarian aid. Remittances, illegal trading and humanitarian assistance all provide ways of recycling foreign exchange so that local actors can purchase weapons and equipment (as well as luxuries) on international markets.

Remittances from abroad have become an increasingly important source of finance for the "new wars". In Kosovo, for example, the shift in diaspora funding from the non-violent movement to the KLA was a key development in the descent towards war in the late 1990s. An organization like Hezbollah in Lebanon depends on support from its members who work in places like the Gulf or Western Europe, as well as from outside powers like Iran and Syria. In Southern Sudan, fighters gained access to remittances sent to poor families from family members working in the Gulf by controlling the price of food at checkpoints.

Another increasingly important source of finance is smuggling in goods like cigarettes, alcohol and drugs as well as human beings (illegal immigrants or abducted women and children) and illegal trading in valuable commodities like oil and diamonds. Countries which are dependent on valuable primary commodities have been shown to be particularly prone to violence (Kaldor, Karl and Said 2007). This is partly because of geopolitical competition for control over the sources of valuable commodities. But it is also because such countries develop rentier societies in which different groups or individuals struggle for access to the state because it means access to oil or diamond revenues and because commodity revenues can be used to finance wars. In Chechnya, local warlords sold oil drawn from backyard oil wells

to Russian generals who, in turn, sold the oil they had received from the government on the Moscow market in order to finance soldiers' wages. In Colombia, different irregular forces have developed different techniques for accessing oil revenues: blowing up pipelines, taking oil workers hostage, controlling municipalities that receive oil revenues. In the war in Angola the government was financed by oil revenues and the rebels by diamond revenues.

In other words, in place of the national formal economy, with its emphasis on industrial production and state regulation, a new type of globalized informal economy is being established in which external flows, especially humanitarian assistance, remittances from abroad, transnational smuggling and trade in valuable commodities, are integrated into a local and regional economy based on looting, pillaging, hostage-taking and forced manipulation of local prices. Some argue that these new types of violence are actually motivated by private gain. It is certainly true that, in conditions of conflict where the rule of law breaks down, there are also many opportunities for criminal activity. In Iraq and Gaza, much hostage-taking is for ransom rather than to achieve political goals. In the former Yugoslavia, many paramilitary fighters were able to take control of new homes or small businesses. But it is also true that many politically motivated fighters or terrorists engage in economic crime in order to finance political violence. Moreover, many of these forms of violence are also criminal in the sense that they violate international law. The point is rather that the difference between organized crime and war is breaking down in our contemporary era.

This new type of predatory political economy has a tendency to spread. The cost of war in terms of lost trade, especially where sanctions or communications blockades are introduced or where borders are closed, either deliberately or because of fighting; the burden of refugees, since it is generally the neighboring states who accept the largest numbers; the spread of illegal circuits of trade; and the spillover of identity politics—all these factors reproduce the conditions that nurture the new forms of violence. It is possible to identify "bad neighborhoods" where weak states are further weakened by the spread of the globalized informal economy—these include places like West and Central Africa and the Horn of Africa; South-East Europe, the Caucasus and Central Asia; and Iraq, Lebanon and Israel/Palestine.

II.4 Establishing Protective Security

> What is the difference between a barroom brawl and a boxing match? Nothing, save that one is recognised as legitimate based on certain fictions (no one will get hurt, etc.).
>
> What is the difference between the Cosa Nostra and state sovereignty? Nothing, save that one is recognised as legitimate, but based equally on a series of fictions.

> This is why violence is kept at arm's length—the carefully crafted notions that war takes place on "battlefields" and that criminal violence is constituted of marginal elements that can be contained . . . the illusion that violence is "outside society" and that the state keeps society, keeps *us* safe.
>
> South African advocate Justin Wylie, in an interview with Carolyn Nordstrom (2004: 80)

I have argued that the illusion that violence is "outside society" and that through war, the state keeps us safe is becoming harder to sustain because war itself produces insecurity, either as a counter-reaction, or because violence is inflicted on civilians rather than enemies, or because of the growing links between political violence and organized crime. If protective security is to be recognized as legitimate, it has to have some basis in reality. The gap between what is carried out in the name of security and people's lived experience has to be narrowed. Given the growing ineffectiveness of military force, the other two aspects of legitimacy—wider assent from the population and external recognition—are becoming more and more important.

The best way to provide protective security is what happens inside some well-run societies—namely a rule of law, backed up both by the consent of the population and by protection services, such as the police, and emergency services, including firefighters and medical services. Rather than appearing to keep the violence "outside", what is needed is to spread the "inside" outside so that every society experiences a similar level of protection. If we are to counter the new forms of violence, we need, first and foremost, to protect civilians—the main victims of the violence. This requires the (re-)establishment of a legitimate political authority that can guarantee a rule of law on the basis of trust, and that can (re-)establish a monopoly of legitimate violence (to deal with the proliferation of non-state actors) and a legitimate economy (to counter transnational criminal activity), and that, above all, uses new communications to transmit a convincing message about justice, fairness and reconciliation so as to counter the propaganda of hate.

But this has to happen within a multilateral framework. Protective security can no longer be provided unilaterally. States do need to provide protective security for their citizens, but within a global framework that constrains those states that produce insecurity either as a consequence of authoritarianism or state weakness or both. The combination of international humanitarian law and human rights law does provide the basis for an emerging international legal regime based on the rights of individuals rather than states. States, of course, remain the only authorities capable of upholding the legitimate use of force but their use of force is much more circumscribed than previously by international rules and norms. States derive their legitimacy from an explicit or implicit social contract in which they provide protective security in exchange for taxation and votes. Not only does the protective security side of the social contract need to be renegotiated, but it also needs to

be supplemented by a global social contract in which the citizens of the world are guaranteed protective security by the club of states within the framework of the United Nations.

There is a role for military forces in such a system, but not for war against other states—rather, as a contribution to global security and to implementing a global social contract which enshrines human rights. Military forces are required, in conjunction with police and other civilian experts, in cases where international law is being flouted and there is an impending humanitarian catastrophe, e.g. ethnic cleansing, genocide or other large-scale violations of human rights. The legitimate use of military force by states would need to be approved by the United Nations or to conform to a clear set of criteria that are agreed internationally. The adoption of a "responsibility to protect" by the United Nations General Assembly in 2005 is an important step in this direction.

Essentially, military forces would be used together with other types of protection forces, for law enforcement rather than for war-fighting or peacekeeping. Both war-fighting and peacekeeping are defined in terms of a war between collective enemies. The job of war-fighting is about defeating enemies and even though counter-insurgency operations often adopt a "hearts and minds" approach, the task of protecting civilians is secondary to the task of defeating the enemy. The job of peacekeepers is to separate warring parties, monitor ceasefires or collect weapons; in the past, peacekeepers have often been unable to prevent violations of human rights. In law enforcement operations, the main job is protecting civilians.

But using military forces in a law enforcement role is very different from a war-fighting or peacekeeping role. The ways in which military force should be used are as important as the ends. In what follows, I briefly define three principles that distinguish the use of military forces in a law enforcement role from their role in war-fighting or peacekeeping.

First of all, the primary task of the law enforcement role is the protection of civilians, upholding human rights. Killing is never permissible except in self-defense or to save a third party. Thus the killing of an attacker is only permissible if it is necessary to save civilian lives. Of course, it could be argued that this also allows for war-fighting actions that may risk civilian lives. The British forces defending the United Nations safe haven of Goradze in the last stages of the war in Bosnia-Herzegovina did shell Serb forces for several hours in order to prevent the Serbs from overrunning the town and in order to negotiate the safe passage of civilians. This was in contrast to Srebrenica, where Dutch forces failed to prevent the massacre of 8,000 men and boys. One of the reasons why mandates were so restricted in Bosnia-Herzegovina was because of fears that active defense of safe havens could slide into war; the term "crossing the Mogadishu line" was coined by General Rose in reference to the disastrous consequences of a shifting to a war-fighting strategy in Somalia. However, there is a difference between active protection and war-fighting. In "new wars", the warring parties do try to avoid battle because the growing

symmetry of military technology makes the outcome dangerous and uncertain. One should not, of course, dismiss the risk of escalation or of an unconstrained extension of violence; in the confusion and emotion surrounding all wars, warring parties do not behave as expected and an extremist logic often takes over, as Clausewitz argued. But the starting point, in ethical and operational terms, is protection rather than defeating an enemy, as opposed to the other way round, which is the characteristic of war.

The second principle, linked to the first, is that protection can be achieved through stabilization rather than victory. The point of stabilization is to create space for a political process that can lead to the establishment of legitimate political authority. As Rupert Smith, the former Commander of UNPROFOR in Bosnia, has put it:

> We do not intervene to take or hold territory.... Instead we intervene in ... order to establish a condition in which the political objective can be achieved by other means and in other ways. We seek to create a conceptual space for diplomacy, economic incentives, political pressure and other measures to create a desired political outcome of stability, and, if possible, democracy. (Smith 2006: 270)

Stabilization can only be achieved with local support and consent. Of course, it can be argued that military victory is an effective method of stabilization; this is what the Americans claim to have done in Iraq and it is a view that runs deep in military establishments. But in some cases, military victory may simply be beyond reach—every excessive use of force further inflames the situation. In other cases, short-term military victory can be achieved but the cost in terms both of casualties and political legitimacy is too high. Israeli forces, for instance, have succeeded in slowing down the rate of suicide bombing but this has not led to any resolution of the conflict; indeed, it has only inflamed passion more highly on the Palestinian side. Military victory may mean that stability can only be sustained through massive repression and coercion. In Algeria, the French won militarily but lost politically and the trauma of that war has left a lasting legacy.

A third principle is that those who violate human rights are individual criminals rather than collective enemies. This means that military forces or the police have the job of arresting criminals and bringing them to justice. It also delegitimizes the enemy, who are no longer political foes but law-breakers (Kaldor and Salmon 2006). Thus British forces operating in Sierra Leone chose to arrest members of the "West Side Boys" engaged in looting and pillaging a village rather than engaging them in a firefight. This greatly diminished their stature and correspondingly raised the credibility of the British forces. This approach, of course, is not easy; there is often a tension between what counts as political and what counts as criminal. In some cases, it is important to outlaw those who have committed terrible crimes in order to establish a legitimate political process. This is the case in the former Yugoslavia, where, in principle, excluding indicted criminals creates space for more moderate

politics. Likewise, the political situation in Iraq would have been greatly improved had there been a legal process, in the initial stages, leading to the indictment of those members of the Ba'ath Party who committed unspeakable crimes, thus preventing arbitrary revenge attacks which contributed to the development of sectarian conflict, and allowing those relatively innocent civil servants or teachers who joined the Party out of fear to continue their work. On the other hand, it may also be important to include groups like the IRA or Hamas, who are viewed as legitimate by some parts of the population, in the political process.

The British experience in Northern Ireland does provide an example of how to conduct such operations, especially after 1974, when the emphasis was placed on police primacy and support to civil power. The military did help to stabilize the situation and eventually create the conditions for a peace agreement. What made Northern Ireland different was the fact that the conflict took place on UK territory. Bombing Belfast was not an option. It could also be argued that the different response of American authorities to the Oklahoma bombing as opposed to 11 September can be explained partly by the fact that this was a domestic rather than an international incident. It is no longer possible, or relevant from the point of view of the victims, to distinguish between foreigners and citizens or between the domestic and the international. Although the state has primary responsibility for dealing with domestic violence, there are external situations where the local state itself is the cause of violence or where it is incapable of dealing with violence, where international forces intervene but with methods that are not so very different from the methods that might be used in a domestic setting. This reflects both the changed sensibilities of society, where concerns about people far away have become more urgent as a result of global communications and transnational communities, and an emerging global social contract whereby the international community recognizes individual rights and not just state rights.

Of course, protective security is more than just protection against violence. It has to involve protection against natural disasters and economic crises as well. It has to concern itself with the conditions that lead to violence, such as unemployment and high crime rates. What I have suggested here is that today's military are often part of a protection racket. And I have proposed ways in which the provision of security, and the role of the military in providing that security, needs to change if states are to protect their citizens effectively and so restore their legitimacy.

References

Amnesty International (2007), "Afghanistan: Civilians Bear Cost of Escalating Insurgent Attacks", available at <http://hrw.org/english/docs/2007/04/16/afghan15688.htm>.

Beck, U. (1992), *Risk Society: Towards a New Modernity* (London: Sage).

Burnham, G., Lafta, R., Doocy, S., and Roberts, L. (2006), "Mortality after the 2003 Invasion of Iraq: A Cross-Sectional Cluster Sample Survey", *The Lancet*, 11 October 2006; available online at <www.thelancet.com/webfiles/images/journals/lancet/s0140673606694919.pdf>.

Centre for Research on Globalization (2007), <www.globalresearch.ca>.

Commission on Human Security (2003), *Human Security Now*, available at <www.humansecurity-chs.org/finalreport/English/FinalReport.pdf>.

Hoffmann, S., and Fidler, D. E. (eds), (1991), *Rousseau on International Relations* (Oxford: Clarendon Press).

Howard, M. (1982), "Reassurance and Deterrence", *Foreign Affairs*, 61: 309–24.

Human Rights Watch (2004), *Darfur in Flames: Atrocities in Western Sudan*, available at <www.hrw.org/reports/2004/sudan0404/index.htm>.

Iraq Coalition Casualty Count (2007a), "Iraq Coalition Casualties", available at <http://icasualties.org/oef>.

—— (2007b), "Operation Enduring Freedom: Coalition Fatalities", available at <http://icasualties.org/oef>.

Kaldor, M. (1991), *The Imaginary War* (Oxford: Basil Blackwell).

—— (1999), *New and Old Wars: Organised Violence in a Global Era* (Cambridge: Polity Press).

—— Karl, Y., and Said, Y. (eds) (2007), *Oil Wars* (London: Pluto Press).

—— and Salmon, A. (2006), "Military Force and European Strategy", *Survival*, 1(48): 19–34.

Keen, D. (2007), *Hidden Functions of the War on Terror* (London: Pluto Press).

Nordstrom, C. (2004), *Shadows of War: Violence, Power and International Profiteering in the Twenty-First Century* (Berkeley and Los Angeles: University of California Press).

Sen, A. K. (1999), *Development as Freedom* (Oxford: Oxford University Press).

Small Arms Survey (2007), *Small Arms Survey 2007: Guns and the City* (Cambridge: Cambridge University Press); available at <www.smallarmssurvey.org/files/sas/publications/yearb2007.html>.

Smith, R. (2006), *The Utility of Force: The Art of War in the Modern World* (London: Allen Lane).

Tilly, C. (1982), *War Making and State Making as Organised Crime*, CRSO Working Paper 256 (Ann Arbor: Center for Research on Social Organization, University of Michigan).

UNDP (2006), *Evaluation of UNDP Support for Conflict Affected Countries*, available at <www.lse.ac.uk/Depts/global/Publications/UNDP_Human%20Security%20Report2006.pdf>.

UN News Centre (2007), "Unable to Visit Darfur, UN Human Rights Council Fact-Finding Team heads to Chad", available at <www.un.org/apps/news/story.asp?NewsID=21586&Cr=sudan&Cr1=darfur>.

CHAPTER 26

DEMOCRACY AND ITS INDIAN PASTS

SUNIL KHILNANI

I. Introduction

In a discipline that on the whole has scant concern for its own history, let alone for history more generally, Amartya Sen is distinguished by his interest in connecting the concerns of the present to the past. A sense of the past characterizes both his work within the discipline of economics and his wider inquiries into our present conceptions of ethics, politics and modern identities. It has given him a perspective on the efficacy, the power, and ultimately the limits of economics—and the need for other forms of knowledge and reasoning to step forward at the point where economics has to get off. There are parallels between his intellectual trajectory and that of his friend, the late Bernard Williams—both masters of the tools of their discipline, who with increasing clarity and force have established the limits of what those tools, and indeed their discipline, can do for us.

The concern for historical perspective, for the expansion of our points of view which this can bring, has enabled Sen to reveal to us both a more imaginative conception of our current values—freedom, secularism, human development—and a broader range of means to pursue them. It sets Amartya Sen apart. An economist at the summit of his profession, he also himself belongs to a tradition of Indian intellectuals and political figures, often not professional historians, who over the last century have interrogated the past to enrich our understanding of future possibilities. Whatever their ideological persuasion, these intellectuals sought to

discover antecedents for their favored idea of the political community, and for their vision of a future India. Tagore and Gandhi extensively discussed the relation between the past and present, and Nehru was almost obsessively fascinated by this relationship. In what respects did India need to break away from an oppressive past, and what continuities might it seek to sustain as it worked to build a just society?

The immediate context for Amartya Sen's own interest in the Indian past has been set by developments in India and abroad over the past two decades. Three of these developments are worth noting: the rise of the Hindutva movement, with its ambition to redefine the Indian nation and its founding values; the emergence of post-modern attempts to dismantle national subjects and stories in favor of more fragmentary and particular ones; and the political currency of crude civilizational stereotypes, with their predictions about the inevitably conflictual outcomes of civilizational encounters. If there is one common theme that joins these disparate developments, it is the idea that values like secularism and democracy, and their associated practices and institutions, are uniquely possessions of the West and Europe—that have had to be transplanted or imposed elsewhere. Thus there is little reason to expect—let alone want—them to succeed in non-Western locations.

In a stream of essays and public interventions during the 1990s and 2000s, whose main themes have been gathered into books such as *The Argumentative Indian* (2005) and *Identity and Violence* (2006), Amartya Sen has developed his dual perspective on India and the world. In the first instance, these writings amount to an impressive defense of the founding values of the Indian republic—secularism, cultural pluralism, democracy. But Sen has sought also to demonstrate that these values, or at the very least recognizable analogs to them, have deep roots in India's past. They long precede the emergence of the modern state during the colonial era, he has argued, and continue to provide a resource of norms and criteria by which to judge the performance of contemporary Indian democracy. Sen sees the Indian past as containing a plurality of traditions and viewpoints: those that stress orthodoxy, spirituality and mysticism have received the greatest attention, and yet modern India is equally heir to intellectual traditions that celebrate dissent, skepticism and disputation. And these latter traditions, he claims, have a direct influence on contemporary Indian democracy—making it possible, sustaining it, and potentially offering criteria by which to improve it. In that sense, Indian democracy is not simply a Western graft: the institutions of parliamentary democracy are certainly a legacy of British rule, but these could succeed because of a more general historical tilt towards the valuing of argument and discussion—and agreement based on this.

The relative success of India's contemporary democracy, unprecedented in its scale, has been quite justly celebrated. The fact that the country has, since 1947, held 14 national elections and many more in its regional states; that dozens of peaceful

changes of government have occurred; that Indians participate in elections at impressive and growing levels; that the country has been able to deal with the practical issues involved in summing, within a few hours, the ballots of around 400 million citizens into a collective choice: all of these achievements support the conclusion that India's political system has succeeded in institutionalizing uncertainty—a crucial criterion of any democratic success. That is a significant achievement. It is, however, a partial one—particularly when seen from the point of view proposed by Amartya Sen.

For elections, and their associated practices and results, do not exhaust our sense of what we wish democracy to be—and do—for us. Sen urges a broader view of democracy that looks beyond the capacity of Indians to operate with and through political parties, elections, and the institutions of a democratic state—a view that focuses on social practices (of reasoning, argument and deliberation), and so gives democracy an important life outside the structures of formal politics associated with the state. "Discussion and argument", Sen has insisted, "are critically important for democracy and public reasoning" (2005: p. xiii). This is an important perspective in the Indian context, where, after all, the representation of political power in the form of the state began much later than in Europe, and is associated with the coming of British colonialism. Sen's move thus enables us to speak of democratic forms and practices even prior to, and apart from, the modern state. India's sixty-year-old constitutional democracy, he suggests, can draw on the historical potential of a millennial tradition of heterodoxy and multivocal exchange. By reinstating the centrality of *voice* to democracy, he reminds us immediately of the classical idea of *isegoria*, and points to a more expansive conception of politics—seeing it rooted, beyond the formal structures of the state, in everyday practices of dialog. Sen might be seen here as reformulating a distinction made by Alexis de Tocqueville—between democracy conceived as a general idea or principle (whose presence one can discern even in the European Middle Ages—and, Sen argues, in many other locations at different points in history) and, more narrowly, as a set of institutions of government that developed in parallel with the rise of the modern state.

While Sen turns back to India's history to identify democratic resources, equally he looks outward, seeking to place Indian political ideas and experience in a global context. This is in part because those ideas have of course developed in contact with other civilizations, above all China and Europe (cf. Goody 1996), and also because today, with India's consolidation as the largest, most unusual democracy in human history, we need to see how its inherited intellectual traditions connect to the current, global challenges of politics. India, Sen insists, should not be set apart as exceptional. It needs rather to be integrated into a global view of the history and future possibilities of democracy and democratic argument.

Sen's view of democracy as rooted in voice and public discussion allows for new possibilities of conceiving of "global democracy". He recognizes the impossibility of

a "global state" or of a single "world government": in that sense, global democracy is a chimera. On the other hand, it is possible to identify forms of what Sen has called "borderless public reasoning"—about international justice and rights—which can reach across the limits of state institutions, and establish the possibility of public debate across national and other boundaries. Such a global debate, in which a wide range of voices could participate, would increase the points of view that require consideration in determining how to act to improve levels of justice, or the register of rights, across the world.

This possibility of a more global conception of democracy today must in turn be rooted in a more global history of the sources of democratic politics. Such a history would track the presence of practices of public deliberation and reasoning in a wide range of historical societies, and not see them as unique to the West. Such an approach to the diverse histories of democracy would have an inspirational value to many outside the West who are struggling to establish democracy in their societies; and it would also have a real political dimension. It would help to resist what Sen has called "civilizational partitioning" and "confinement"—a viewpoint that both denies civilizations their internal diversity, and freezes them in a structure of unequal power relations.

II. India's Democracy and its Past Traditions

In the fraught global political climate of the early twenty-first century, Sen's humanistic approach—his willingness to take history in the round, to ask large questions and see large patterns, and his expansion of our sense of democracy beyond both the Western inheritance and the modern state—is profoundly compelling. Equally convincing is his insistence that we cannot have adequate understandings of what democracy is and means today unless we have a richer understanding of its non-Western experiences—among which India's stands out in particular.

His perspective also raises two large sets of questions. First, how should we think about the relationship between India's past history and its present politics? How exactly is the persistence of India's democracy, which defies classical theoretical precepts, related to the long history of the country's political ideas and institutions? Beyond the usual dichotomies associated with imperialist or nationalist positions, which see Indian democracy as simply a legacy of the British Raj or, contrarily, as rooted in age-old village republics, does a focus on what Sen sees as India's historical traditions of dialog and discussion help to *explain* the character of India's contemporary democracy? Or is Sen's purpose somewhat different: namely,

to remind us of these traditions of disputation and dialog, show us some of their achievements, and thereby provide us with criteria by which to assess and improve India's actually existing democracy?

Second, how can we pursue the task of conceiving the history of democracy and democratic politics within a more globalized framework—one that can expand on and modify the tenaciously held perspective that democracy is purely an achievement of Western history? And what would be the place, within this more globalized view, of India's democracy—which, each time it holds a national election, is engaged in the largest such act of free choice in human history?

This twin intellectual project—namely, to devise an understanding of India's democracy that integrates it into the country's own deep history, and in turn to locate this understanding within a global view of democracy—suggests the ambition and challenge of Sen's writings on India's past and its current democracy. It is a task that we are still far from meeting. It pursuing it, however, there are several difficulties, and I would like now to consider three of them, beginning with some brief remarks about the range of materials out of which to construct a history of political ideas in India. Second, I shall suggest some difficulties surrounding the idea of a continuous Indian "argumentative tradition"—and particularly, how we might assess the significance of this idea for India's modern democracy. And finally, I shall raise some difficulties with the conception of democracy that Sen deploys.

Sen's claims about India's intellectual heritage provokes the following question: what is a useful, appropriate way to conceive the long-term historical relationship between the intellectual traditions and the political ideas and practices that have created contemporary India? Is there in fact any meaningful connection between what Sen has called "the long history and consummate strength of our argumentative tradition" (2005: p. vii), and the practices and possibilities of Indian democracy today? He is of course fully aware of the range of opposed, non-democratic resources which are also available in Indian traditions—the deep potentials that run in the opposite direction (his point is exactly that strands of orthodoxy and heterodoxy are in argument with one another).

Sen quite explicitly and self-consciously sees his alternative genealogy of the Indian past, which lays emphasis on the presence of heterodoxy and multivocal discussion in India's high Sanskrit, Buddhist and Mughal past, as highly selective. That selectivity has drawn the ire of critics who, while sharing his liberal values, remain skeptical of the soundness of his lineage of precedents for India's modern, electoral democracy. Sen, however, has a rationale and defense for his approach—which he has demonstrated in a number of significant pieces on the methodology of the social sciences (Sen 1981, 1982, 1993). Our interest in the social or historical world is driven by practical reason. The kind of knowledge we want is the kind that will be in some way useful to us, whatever our purposes happen to be. The ambition of an absolutist conception of knowledge, the view from nowhere, is simply an inhuman one: all views are views from somewhere. But it does not follow

that all views are equally valid. What constrains the perspectives of what Sen calls "positional objectivity" is a commitment to epistemic standards of evidence and truth—themselves constrained by criteria of relevance.

Sen's interest in the past is not, of course, primarily that of an historian—it is purposefully partial, designed to bend the stick the other way, to counter what he sees as a one-sided view of India's "real heritage" as purely a religious one. He has an explicit, justified criterion for selection—to suggest the extent of India's intellectual heterodoxy. But one might still note some limitations with his selectivity, particularly if we wish to understand more generally the history of political ideas and practice in India.

Recent scholarship on the pre-colonial Indian languages of politics and good government has widened the scope of materials that merit consideration. The conventional accounts of Indian political ideas customarily begin with Kautilya's *Arthashastra* (a text about whose context and reception in fact we know rather little), and then leap over several thousand years to Ram Mohun Roy, the founder of modern Indian political thought. It is as if one told the history of European political thought by beginning with, say, Plato or Aristotle and then leapt straight to Bentham or Paine. It remains difficult, within this perspective, to see continuities in anything except very broad terms. In effect, we are given a choice: either we accept very general connections between contemporary political practice and ideas and pre-colonial traditions; or we see the entry of colonialism as having produced a total rupture, replacing earlier spiritual and religious traditions with Enlightenment values, which contemporary Indian democratic life seeks to approximate.

The more precise historical reconstruction that is required in order to embed India's modern democracy within its longer historical and intellectual lineages is still in its very early stages. We know too little about the varieties of political thinking and debate in India: we need still to establish the linguistic identity of these ideas (whether in Sanskritic, Persian or vernacular traditions), the genres in which they were expressed, the audiences to whom they were directed—and their actual social effects. And without a fuller picture, it is hard to assess the dialogic or democratic potentials in India's past, or to gauge their relative importance and effects.

One significant bias in our understanding has been excessive attention to the high Sanskritic traditions (as embodied for instance in the Dharmashastric texts). This bias towards examples from Sanskrit traditions is evident in Sen's own work, which stresses the "quintessential 'Indianness'" of Sanskrit (2005: 82). Yet, as is becoming apparent from the work of scholars on Persian sources as well as vernacular languages, there may be significant resources in India's other linguistic cultures to support ideas of moderated power and good political judgement (Alam 2004; Subrahmanyan and Narayana Rao 2007). Some of this work in fact shares some of Sen's general corrective intentions, originating, for instance, in studies of southern

Indian practices of historiographical and history writing, which were designed to show that, contrary to British insistence, there are developed conceptions of history and historical time within India's traditions (just as Sen wishes to show the presence of dialogic and deliberative elements in classical and pre-colonial India). Extending their earlier work, which was designed to show the availability of a concept of history in "early modern India", Subrahmanyam and Rao have argued that conceptions of politics can also be found in the vernacular traditions of the Deccan and of southern India, conceptions that are distinct from those found in the Sanskritic traditions. They and others have identified a line of *Niti* texts in regional languages (especially Telegu and Marathi), produced in the early modern period. These texts offered practical precepts and guidance in political matters, and dispensed moral advice (about, for instance, moderation in power or cultivating good judgement) which contrasted with the social conservatism of the Dharmashastric texts. And they were consciously "this-worldly" or secular in character. The Dalit leader and constitutionalist B. R. Ambedkar himself drew a contrast between what he called the "secular law" of the *Niti* tradition and the "ecclesiastical law" of the dharmic tradition—which suggests a connection with modern Indian secularism worth exploring. Such work (and there are other examples) is beginning to trace historical filiations and continuities between past and current politics.

Beyond the actual content to be found in India's diverse linguistic cultures and its relevance to ideas of moderate and consensual power, there is also a larger, theoretical point at stake. Arguably, it is the very fact of India's linguistic multiplicity that has helped to foster positive attitudes towards pluralism and diversity—both within and between different linguistic cultures. This second-order diversity has itself encouraged heterodoxy and the "dialogic heritage" to which Sen has drawn our attention (2005: 44 *inter alia*). As Sheldon Pollock has demonstrated in his remarkable study of the cultural and political role of Sanskrit in pre-modern India, one of the notable features of India's cultural pluralism is that political power did not seek to mobilize linguistic difference and particularism for its own purposes: "No language in southern Asia ever became the target of direct royal regulation; sanctions were never imposed requiring the use of one and prohibiting the use of another", Pollock writes; "At the time when episodes of vernacular extermination were occurring in Europe, kings in Karnataka were issuing royal records in Kannada for the core of their culture-power *desa*, in Telugu for the eastern sector, and in Marathi for the western, and in their courts these kings were entertained with songs in these languages as well as Avadhi, Bihari, Bangla, Oriya, and Madhyadeshiya—producing, in fact, a virtual cosmopolitanism of the vernaculars" (Pollock 2006: 573). This developed historical capacity for living with linguistic heterogeneity has encouraged the tolerance of cultural diversity—which in turn has been a vital condition for the possibility of Indian democracy.

III. Tradition, Contexts, and Meanings of Argument

Yet, even if we did have an adequate historical map of the histories of India's classical and vernacular political languages, there would still remain questions about whether it is possible to identify a "long history of the argumentative tradition in India", and about the conception of democracy which Sen upholds.

Beyond the quite substantial familiarity that Indians now have with electoral democracy, Sen has reminded them that they also have other inheritances which can support and expand their democratic practice. Indians are heir to argumentative resources, embodied in philosophical and literary traditions, which can support reasoned debate—about the ends of democracy, about what social values to promote. Such debate is vital to enhancing and improving the practice of Indian democracy, and it serves to expand the range of voices and points of view that define India's democracy—and the goals it chooses to pursue.

But in what sense can one speak of an argumentative *tradition* in India? The examples and episodes that Sen collects do testify to the presence of skeptical, rationalist and critical currents in Indian thought. But do they collectively yield a continuous line of reflection? Or are they in fact simply sporadic episodes, connected retrospectively through the lens of a liberal sensibility—Sen's own?

The examples Sen cites—among others, Ashoka, Lokayatta disputations, Kalidas, Akbar—have indeed had a persisting imaginative presence in India. During the twentieth century, for instance, India's intellectual elites quite often also invoked these examples and the norms they suggest, in order to legitimate their vision of India as a pluralist, tolerant culture. Yet these examples cannot be said to have sustained a continuous tradition, nor to have diffused the effects of such examples across a wide social space (cf. Guha 2005). Unlike, say, the idea of the Ancient Constitution in seventeenth-century England, which both propagated and relied upon a sense of England as an historically continuous entity, with a persisting legal tradition—an idea whose reach was widely felt across society (Pocock 1957)—the examples drawn from the Indian case represent at best little oases, isolated pockets in what otherwise is the history of a society that kept democratic and dialogic options well suppressed. They are oases in two senses: they have cropped up sporadically over time and also within their societies, restricted to small circles.

Can we at least speak of them as having yielded normative concepts or principles which have carried a stable meaning across Indian history? Even here one has to be circumspect. Such episodes of argumentation neither initiated nor sustained a continuous, self-conscious reflection on the values of justice, equality or liberty. A tradition implies some notion of self-reflection and self-consciousness: a sense of itself as a tradition, a cumulative chain of thought and practice. Bernard Williams,

in his studies of Homeric literature, could argue for continuity between the ethical ideas of ancient Greece and modern Europe: "we are linked to the past by a self-conscious set of traditions. We reach the past by a route each step of which, in the favourable case, stands in explanatory relations to the steps before and after it" (2006: 61). But it is not clear that we can speak in a similar way of an Indian "tradition", or traditions, of heterodoxy and public discussion. Such a notion invites several questions—not least, in what sense can we refer to a "tradition" of argument or public reasoning? Does the mere sporadic presence in Indian history of certain episodes which, from the point of view of liberal values, look attractive and recognizable in liberal terms, constitute a "tradition"—one which is in genuine need of historical recovery? Or is the term "tradition" being used more loosely?

A further set of questions turns on the role and effects of these episodes in India's intellectual history. Why has this manner of thought, strong as it might have been, remained so ineffectual in actually changing anything in the Indian social order? Sen draws a picture of widespread debate and argument both within and outside religious circles—among exponents of Buddhism, but also among scholars of the Lokayatta, Carvaka and Nyaya philosophies. Yet we have little sense from his account of the ends and purposes of such arguments. All argument is motivated, driven and shaped by interests. Much fuller exploration is needed of *why* there was such a level of debate within India. Were these, in general, largely formalist arguments—scholastic disputations with an internal focus, and not intended to grasp the social world or directed at changing it? Sen does claim, for instance, that the Buddhist councils, held between the fifth century BCE and the second century CE, displayed "a commitment to discussion as a means of social progress" (2005: 15). And yet on the whole these instances of public deliberation left very little mark on the wider society. Max Weber, in his sociology of Indian religions, had also noted the presence within Hinduism of highly sophisticated rationalist traditions. But he also pointed out that somehow these remained unconnected to any projects of social transformation, or to systematic inquiry about the natural world (Weber 1958). Indeed, one might draw an analogy between the effects of the caste system and the examples of intellectual heterodoxy Sen invokes: just as the caste system was able to accommodate many forms of social difference, and integrate them into an order of immense stability, so one might see the acceptance of intellectual heterodoxy as a way of disarming practical criticism. In India, intellectual heterodoxy seems to have long co-existed with great social orthopraxy (Kaviraj 2002). One is reminded here of Herbert Marcuse's once-famous thesis of "repressive tolerance"—a peculiarly Californian condition, but perhaps also one found in pre-modern India.

By clarifying the various motivations of these arguments—and how these might explain their variable effects—we can better understand how insulated these debates and arguments were from the wider social order. They neither developed nor fed into a wider belief that thought, intellection and public reasoning could actually effect change in the world. That view can be contrasted with examples

drawn from the twentieth century. Gandhi's extraordinary daily dialog and argument with fellow Indians (and of course also with the British and others), through the medium of his newspapers, weeklies and correspondence, was a form of argument clearly designed to achieve change. So too were the debates during what is perhaps India's most remarkable moment of deliberation and discussion—the constitutional debates of the Constituent Assembly in the mid-twentieth century. These debates represent the most focused, motivated episode of public reasoning in Indian history, quite explicitly aimed at remaking the social and political world—by breaking with the past (Austin 1966). By comparison, it is hard to detect anything like a similar will to change the world in the earlier forms of public reasoning noted in *The Argumentative Indian*. Here one might in fact question the virtue of continuity, and the effects of "long traditions". For it may be that a self-conscious rupture with the past is actually necessary to redirect argument towards action.

If the notion of tradition itself poses questions, ambiguities also surround the concept of "argument". While many advocates of the virtues of democracy have emphasized the aspect of public debate, discussion, and argument, what exactly is meant by such terms varies. One might for instance distinguish between disinterested discussion and interested argument—where the latter is more like haggling or bargaining (Manin 1997). Or again one might distinguish between argument which aims at an all-or-nothing outcome, which seeks a kind of full-scale conversion to another point of view, and argument which seeks more partial success; or, temporally, between argument which seeks to justify a course of action already performed, and argument which seeks to persuade others to do something. The meanings of arguments are also defined by the types of audience to whom they are directed. Interlocutors may direct arguments at each other, or towards a third party (as in a courtroom)—and seek to convince this third party, rather than each other, of their cause. That is to say, argument may be designed to encourage impartiality, as against adversarial negotiation—and this will often depend to a large extent on the context and institutional structures within which the argument takes place. (In a different context, Sen 2006c draws a distinction between the model of "fair arbitration" associated with Adam Smith's idea of the impartial observer, and Rawls's contractarian idea of "fair negotiation".) Arguments might be made in order to exchange information or to make it public. And arguments may appeal to self-interest or to universal principles, they may invoke threats of immediate sanction as against long-term consequences, and so on.

Different types of argument will stand in differing relations to democratic politics. Notions of deliberation and public reasoning often appeal to more universal, non-contextual principles. But political argument is characterized by its heavily contextual character: it is often motivated by intensely local interests, and appeals to others on the basis of local interests. In a brilliant essay, the literary scholar and poet A. K. Ramanujan described the "inexorable contextuality" of Indian thought—the

fundamental role within it of contextual reasoning and argument (Ramanujan 1989: 54). Paraphrasing Ramanujan's question, one might ask: "Is there an Indian way of arguing?" And one might have to conclude that discussion in Indian democratic politics today more readily approximates to haggling than to appealing to rational principles. In fact, one might want to say (if trying to identify an Indian "national character") that the great talent of Indians is not so much argument, but rather *compromise*: the ability to share the spoils, moral or material. Instead of the "Argumentative Indian", one might speak of the "Compromising Indian".

IV. Democracy and its Content: The Sovereignty and Limits of Reasoning

"Important as history is," Sen has written, "reasoning has to go beyond the past" (2005: 120). Authority, for Sen, lies of course not in tradition or the past, but in reason and the conclusions which reasoning supports. So let us put aside questions of whether historical traditions might sustain democracy in India. Sen asserts the "sovereignty of reasoning"—reasoning unbound by the legacy of past traditions, and which reigns presumably over the possibilities for democracy as well. On this view, it is public reasoning which must scrutinize and assess democracy wherever it is practiced as an institutional form, and which must (wherever it can) assign democracy its content. There is, it would seem, some circularity here. Public reasoning requires the condition of free inquiry, which democracy sustains; and in turn, public reasoning enables democracy to pursue and achieve optimum outcomes. What is the conception of democracy, and its relationship to rationality, that seems to be implied by Sen's account?

An enduring tension within the idea of democracy is that between the authority ascribed to number, and the claims of reason. This is of course a far from purely theoretical tension. Indeed, it has animated Indian politics over the past 60 years, and has if anything sharpened as India's democracy has expanded to an unimaginable scale, drawing in hundreds of millions of citizens, the vast majority of them without significant formal education. This tension is also dramatized in the political career of the single individual who perhaps did most to establish a space for public reasoning and democracy in independent India: Jawaharlal Nehru. Nehru certainly had some imaginative impact on the young Amartya Sen (a copy of Nehru's *Discovery of India* was among the small number of books Sen was able to take with him when he first left India for Cambridge). The case of Nehru is interesting from this point of view because he did have a strong sense of what the appropriate content of democratic choice should be (he certainly had his own

conception of the public good), and yet was also committed to respect the authority that democracy assigns to numbers (Khilnani 2007).

Sen himself has devoted a great deal of thought to the unstable relationship between rationality and democracy. What have been termed the aggregative and deliberative strands of democratic theory are both present in his thinking about democracy (they relate to his distinction, mentioned above, between democracy as an idea or principle, and as a set of institutions and a form of government). The aggregative perspective sees democracy as a problem in social choice (is there a rational collective ordering of myriad individual choices?), as against one that sees democracy as a form of collective deliberation and discussion for the proposal and selection of justifiable choices. Sen's own career was launched by his work on the central problem in the field of social choice theory, the aggregative dimension of democratic decision-making; and so far his most enduring theoretical contributions have been in this field. In recent years, exponents of social choice and public choice theory have been increasingly concerned with the content of choices—the processes whereby choices and values are formed, and the role of reasoning in this—which has led them to a view of democracy as "government by discussion". Sen himself has extended his analysis of democracy to more detailed consideration of the deliberative aspects.

What links these two dimensions of democracy in Sen's own imagination is his abiding concern with choice itself. It is this aspect of human existence that seems to fascinate him most, both at the individual and social levels. Perhaps uniquely amongst all creatures, humans make reflective choices. How are we enabled to choose? What are the bases of our choices? What impediments stand in the way of making them? Above all, Sen is interested in the rationality of our choices—both in the sense of what sorts of reasons we give for our choices, and to what extent our individual choices can yield collectively rational outcomes.

Yet, one of the oldest and deepest criticisms of democracy is that it is a profoundly unreliable method for arriving at well-judged claims or views about the public good—for identifying choices which really are worthwhile to pursue. Rhetoric, demagoguery, caprice, weakness of will, the grip of self-interest: all are much more likely to prevail in public decision-making processes than the claims of rationality. If deliberation is valued because it delivers good epistemic outcomes, then democratic deliberation is a very indeterminate and unlikely method to achieve this. Democracy, understood as the dispersion of collective judgement across a citizen body (that is, the abolition of any division of labor in the production of political rationality), has the clear advantage over any other system of minimizing the exclusion of voices. But this offers no assurances about the formation of well-judged choices—and this is the aspect of democracy that has attracted the most skepticism and adverse criticism, even from those who are sympathetic to democracy.

There have been two (linked) strategies of response to such skepticism. One has been to set out elaborate formal procedures of public argumentation and reasoning,

designed to excise all irrational and self-interested considerations, and to discount the operations of power, and thereby insure rational content to public deliberation: the method of Rawls, with his veil of ignorance, or Habermas and his ideal speech situation. A second, more classical strategy has been to hope for the diffusion of rationality across society through public instruction. From Condorcet via John Stuart Mill and on to John Dewey, education has been the favored instrument for achieving this effect. Yet democratic theory has not (and cannot) devise any reliable way to ensure that its deliberative practices yield determinate cognitive outcomes.

Consider Sen's example of the Mughal emperor Akbar, and his commitment to rational discussion of different religious viewpoints. What attracts Sen to Akbar is that Akbar's practices chime with Sen's own liberal values. But Akbar of course was hardly a democrat, let alone a liberal. The implied converse of this equally holds: there is nothing about democracy itself which can guarantee the production of the values of liberal pluralism.

As India's recent political history bears out, electoral democracy can result in governments that have little interest in, or taste for, liberal values. This was the case with the Bharatiya Janata Party (BJP)—and contrary to what Sen has suggested (2005: 70–1), it is far from clear that the BJP was defeated in 2004 either because of a nationwide rejection of their values, or as a result of a more enlightened understanding among the populace of the values of pluralism and secularism. The 2004 electoral outcome was in fact a national verdict composed of a series of local electoral results, reflecting particular issues and contests. It is above all the institutional structure of Indian democracy, a complex federal political system, which makes it difficult for monolithic ideologies to dominate for long spells at the national level. This same system, however, can allow such ideologies to dominate India's regional states—the BJP, in a particularly extreme form, has dominated the politics of the state of Gujarat for over a decade now. The BJP is perfectly likely to be elected nationally again at some point. There are no democratic safeguards against that possibility; nor is there anything in the Indian past that might help—for example, traditions of pluralism or heterodoxy (not least because, as Sen himself recognizes, there are plenty of counter-traditions that can also be invoked). The phenomenon of the BJP, and its electoral successes, do pose questions for Sen's conception of democracy. He has sought to explain the exclusivist politics of a party like the BJP as essentially an intellectual error—the result of "obscurantism", propaganda and lack of education. Yet, some of the strongest and most extensive support for the BJP comes from quite highly educated social groups, in India and abroad. The political appeal of parties like the BJP and of the broader movement associated with it, the Sangh Parivar, and the complex psychological chains of identification which link its supporters to the party and its leaders, cannot necessarily be dissolved by the focused beams of rationality.

Amartya Sen has of course reflected deeply and with profound insight on the persistence of intolerance and the violence often associated with it: from his direct, bewildering encounter as an 11-year-old boy with the murderousness of religious

identities during India's partition, to his most recent work on the perils of "solitarist" understandings of identity (Sen 2006), he has championed a more supple conception both of the individual self as a choosing agent, and of our cultural and civilizational affiliations. His instrument has been reason, a faith in public deliberation and in reasoning; and when the full flare of his own mind is on display, problems do seem to burn away. Sen has always sought to increase the space for reasoned discussion and debate, resisting both theoretical absolutisms and "impossibility" theorems, as well as political absolutisms and civilizational imperialisms.

The special potency of democracy, though, lies not in its virtues as a cognitive method for producing public reason, but in its unmatched capacity to delegitimate existing structures of power: its ability to deauthorize regimes, to throw governments out. The relationship between democracy and public deliberation is a tenuous one. It rests largely on the classical idea of *isegoria*: not merely the right of all to enter into political deliberation, but the actual effective opportunity to do so. Yet in modern democracies, there is not in fact any such equalization of power. Modern democracies do recognize the normative entitlement of all to participate, and the cognitive efficiency of bringing into consideration a greater range of voices (and so reducing the danger of excluding relevant concerns) in deciding about a particular issue. This normative claim lends legitimacy to democracy; but in practice, no democracy does (or can) achieve anything of the kind, given the political division of labor that exists. Modern democracy, as John Dunn has insisted, "does not envisage the structuring of public deliberation as a causal field, but instead dissolves it into an aspect of political authorization" (Dunn 2007). It therefore cannot be relied on to produce any particular determinate positive outcomes.

Democracy displaces the sovereignty of reasoning, and installs politics in its place—which necessarily implies a complex moral psychology, in which, apart from reason, there are many other elements that struggle for dominance. Democracy can be seen as a recognition of the limits of reasoning—and in that sense a more appropriate conception of the space and limits within which humans have to live their lives. The achievement of a democracy, the test of its durability, does not ultimately lie in its ability to deal with the rational claims and arguments that are produced within it. It lies in its ability to deal with, to absorb, the claims of the irrational, the self-interested—those who for many complex reasons (fear, despair, cruelty) refuse the injunctions of public reasoning.

References

Alam, Muzzafar (2004), *The Languages of Political Islam in India, c.1200–1800* (Delhi: Permanent Black).

Austin, Granville (1966), *The Indian Constitution: Cornerstone of a Nation* (Oxford: Oxford University Press).

Dunn, John (2007), "Capitalist Democracy or Beguiling Illusion?", *Daedulus*, 136(3): 5–13.
Goody, Jack (1996), *The East in the West* (Cambridge: Cambridge University Press).
Guha, Ramachandra (2005), "Arguments with Sen: Arguments about India", *Economic and Political Weekly*, 40(41): 4420–5.
Kaviraj, Sudipta (2002), "Ideas of Freedom in Modern India", in Robert H. Taylor (ed.), *The Idea of Freedom in Africa and Asia* (Stanford, Calif.: Stanford University Press).
Khilnani, Sunil (2007), "The Discovery of Indians: Nehru and Political Possibility", unpubl. Kingsley Martin Memorial Lecture, Cambridge.
Manin, Bernard (1997), *The Principles of Representative Government* (Cambridge: Cambridge University Press).
Pocock, J. G. A. (1957), *The Ancient Constitution and the Feudal Law: A Study of English Historical Thought in the Seventeenth Century* (Cambridge: Cambridge University Press).
Pollock, Sheldon (2006), *The Language of the Gods in the World of Men: Sanskrit, Culture and Power in Premodern India* (Berkeley, Calif.: University of California Press).
Ramanujan, A. K. (1989), "Is There an Indian Way of Thinking? An Informal Essay", *Contributions to Indian Sociology*, n.s., 23(1): 41–58.
Sen, Amartya (1981), "Accounts, Actions and Values: Objectivity of Social Science", in Christopher Lloyd (ed.), *Social Theory and Political Practice* (Oxford: Clarendon Press).
—— (1982), "Description as Choice", in Amartya Sen, *Choice, Welfare and Measurement* (Oxford: Basil Blackwell).
—— (1993), "Positional Objectivity", *Philosophy and Public Affairs*, 22: 126–45.
—— (2005), *The Argumentative Indian: Writings on Indian History, Culture and Identity* (London: Penguin/Allen Lane).
—— (2006a), *Identity and Violence: The Illusion of Destiny* (New York: Norton).
—— (2006b), "Our Past and our Present", *Economic and Political Weekly*, 41(47): 4877–86.
—— (2006c). "What Do We Want from a Theory of Justice?", *Journal of Philosophy*, 103: 215–38.
Subrahmanyam, Sanjay, and Narayana Rao, Velecheru (2007), "History and Politics in the Vernacular: Reflections on the Medieval and Early Modern in South India", unpubl. paper, UCLA and University of Wisconsin-Madison.
Weber, Max, (1958), *The Religion of India: The Sociology of Hinduism and Buddhism* (Glencoe, Ill.: Free Press).
Williams, Bernard (2006), *The Sense of the Past: Essays in the History of Philosophy* (Princeton, NJ: Princeton University Press).

CHAPTER 27

THE CLASH WITHIN

DEMOCRACY AND THE HINDU RIGHT

MARTHA C. NUSSBAUM

I. Democracy Under Siege

On 27 February 2002, the Sabarmati express train arrived in the station of Godhra, in the state of Gujarat, bearing a large group of Hindu pilgrims who were returning from a pilgrimage to the alleged birthplace of the god Rama at Ayodhya (where some years earlier, angry Hindu mobs had destroyed the Babri mosque, which they claim is on top of the remains of Rama's birthplace). The pilgrimage, like many others in recent times, aimed at forcibly constructing a temple over the disputed site, and the mood of the returning passengers, frustrated in their aims by the government and the courts, was angrily emotional. When the train stopped at the station, passengers got into arguments with Muslim vendors and passengers. At least one Muslim vendor was beaten up when he refused to say "Jai Sri Ram"

This chapter contains material drawn from various parts of Nussbaum (2007a), but attempts an overall synthesis that is nowhere presented as such in the book. The book, of course, contains much more detailed discussion of all the figures and issues mentioned here, as well as extensive references to the literature, and interview material.

("Hail Ram"), and a young Muslim girl narrowly escaped forcible abduction. As the train left the station, stones were thrown at it, apparently by Muslims.

Fifteen minutes later, one car of the train erupted in flames. Fifty-eight men, women and children died in the fire. Most of the dead were Hindus. Because the area adjacent to the tracks was an area of Muslim dwellings, and because a Muslim mob had gathered in the region to protest the treatment of Muslims on the train platform, blame was immediately put on Muslims. Many people were arrested, and some of these are still in detention without charge—despite the fact that two independent inquiries into the event have established through careful sifting of the forensic evidence that the fire was most probably a tragic accident, caused by combustion from cooking stoves carried on by passengers and stored under the seats of the train.

In the days that followed, wave upon wave of violence swept through the state. The attackers were Hindus, many of them highly politicized, shouting Hindu right slogans, such as "Jai Sri Ram" (a religious invocation wrenched from its original devotional and peaceful context) and "Jai Sri Hanuman" (a monkey god portrayed by the right as aggressive), along with "Kill, Destroy!" and "Slaughter!" There is copious evidence that the violent retaliation was planned by Hindu extremist organizations before the precipitating event. No one was spared: young children were burned along with their families. Particularly striking was the number of women who were raped, mutilated, in some cases tortured with large metal objects, and then set on fire. Over the course of several weeks, approximately two thousand Muslims were killed. Approximately half of the dead were women, many of whom were raped and tortured before being killed and burned. Children were killed with their parents; fetuses were ripped from the bellies of pregnant women to be tossed into the fire.

Most alarming was the total breakdown in the rule of law—not only at the local level but also at that of state and national government. Police were ordered not to stop the violence. Some egged it on. Gujarat's Chief Minister, Narendra Modi, rationalized and even encouraged the murders. He was later re-elected on a platform that focused on religious hatred. (Because evidence of his criminal activity is so overwhelming, he has been denied a visa to enter the United States.) Meanwhile, the national government showed a culpable indifference. Prime Minister Atal Bihari Vajpayee suggested that religious riots were inevitable wherever Muslims live alongside Hindus, and that troublemaking Muslims must have been to blame. Leading politicians conveyed the message that government would treat the nation's citizens unequally: some would receive the full protection of the law, and others would not. Prosecutions resulting from the riots have faced related problems: the bias of local judges, the intimidation and bribery of witnesses, and so forth.

While Americans have focused on the war on terror, Iraq, and the Middle East, democracy has been under siege in another part of the world. India—the most populous of all democracies, and a country whose constitution protects human

rights even more comprehensively than our own—has been in crisis. Until the spring of 2004, its parliamentary government was increasingly controlled by right-wing Hindu extremists who condone and in some cases actively support violence against minorities, especially the Muslim minority. Many seek fundamental changes in India's pluralistic democracy. Despite their electoral loss that year, these political groups and the social organizations allied with them remain extremely powerful. Democracy and the rule of law have shown impressive strength and resilience, but the future is unclear.

What has been happening in India is a serious threat to the future of democracy in the world. The fact that it has yet to make it onto the radar screen of most Americans is evidence of the way in which terrorism and the war on Iraq have distracted Americans from events and issues of fundamental significance.[1] If we really want to understand the impact of religious nationalism on democratic values, India currently provides a deeply troubling example, and one without which any understanding of the more general phenomenon is dangerously incomplete. It also provides an example of how democracy can survive the assault of religious extremism, from which all modern democracies can learn. In May 2004, the voters of India went to the polls in large numbers. Contrary to all predictions, they gave the Hindu right a resounding defeat. Because even exit polls, taken in cities and towns, did not predict the result, it is clear that impoverished rural voters played a major role in giving India a new government.

In this chapter I shall use the case of Gujarat as a lens through which to conduct a critical examination of the influential thesis of the "clash of civilizations", made famous by Samuel Huntington (1996). Huntington's picture of the world as riven between democratic Western values and an aggressive Muslim monolith does nothing to help us understand today's India, where the violent values of the Hindu right are imports from European fascism of the 1930s, and where the third-largest Muslim population in the world lives in peace, despite severe poverty and other inequalities. I shall argue, through a study of this case, that the real "clash of civilizations" is not the clash *between* "Islam" and "the West", but, instead, a clash *within* virtually all modern nations—between people who are prepared to live with others who are different, on terms of equal respect, and those who seek the protection of homogeneity, and the domination of a single "pure" religious and ethnic tradition. At a deeper level, the thesis of this essay is the Gandhian claim that the real "clash of civilizations" is a clash within the individual self, between the urge to dominate and defile the other and a willingness to live respectfully on terms of compassion and equality, with all the vulnerability that such a life entails.

[1] In analyzing the reaction of the US government, one would also want to consider the importance, at the time, of the idea of a war on Islamic terror; nonetheless, it remains true that the State Department documented the riots accurately in its report on religious freedom, and subsequently refused to grant a visa to Narendra Modi. See Nussbaum (2007a: ch. 1).

This argument about India will also suggest a way to see America, which is also torn between two different pictures of itself. One picture shows America and Americans as good and pure, its enemies as an external "axis of evil". The other picture, the fruit of internal self-criticism, shows America to itself as complex and flawed, torn between forces bent on control and hierarchy and forces that promote democratic equality. At a deeper level, what I've called the Gandhian level, my "internal clash" picture shows Americans to themselves as people each of whom is capable of both respect and aggression, of both democratic mutuality and anxious domination. As George Kennan wrote:

> I wish I could believe that the human impulses which give rise to the nightmares of totalitarianism were ones which Providence had allocated only to other peoples and to which the American people had been graciously left immune. Unfortunately, I know this is not true . . . The fact of the matter is that there is a little bit of the totalitarian buried somewhere, way down deep, in each and every one of us. (1990: 168)

My argument, then, is focused on India, but it is also pertinent to other countries: for, as Nehru said on the eve of India's independence, "all the nations and peoples are too closely knit together today for any one of them to imagine that it can live apart".

II. The Ideology of the Hindu Right

According to the Huntington thesis, each "civilization" has its own, rather unitary, distinctive view of life, and Hinduism counts as a distinct "civilization". If we investigate the history of the Hindu right, however, we will see a very different story: traditional Hinduism was decentralized, plural and highly tolerant, so much so that the vision of a unitary "pure" Hinduism that could provide the new nation with an aggressive ideology of homogeneity could not be found in India at all: the founders of the Hindu right had to import it from Europe. Today, European fascism is seated right at the heart of what parades, in some quarters at least, as Hindu civilization.

The Hindu right's view of history is a simple one; like all simple tales, it is largely a fabrication, but its importance to the movement may be seen by the intensity with which its members go after scholars who present a more nuanced and accurate view: not only strident public critiques, but organized campaigns of threat and intimidation, culminating in some cases in physical violence. Here's how the story goes.

Once there lived in the Indus Valley a pure and peaceful people. They spoke Vedic Sanskrit, a language revealed as that of the gods when the immortal Vedas were given to humanity. They had a rich material culture, well suited to sustain

their prosperous life. Despite their peaceful temper, they were also well prepared for war: they had chariots, and even horses. Their realm was vast, stretching from Kashmir in the north to Sri Lanka (Ceylon) in the south. And yet they saw unity and solidarity in their shared ways of life, calling themselves Hindus and their land Hindustan. No class divisions troubled them; nor was caste a painful source of division. The condition of women was excellent.

This peaceful life went on for centuries. Although from time to time marauders made their appearance at this people's doorstep (for example the Huns), they were quickly dispatched, because this people was aggressive when it needed to be, and its warlike strength was feared far and wide. Suddenly, rudely, unprovoked, invading Muslims put an end to all that. The early medieval period saw brief incursions by Muslims bent on the destruction of Hindu temples; these, however, proved short-lived. Disaster struck with a heavier hand, however, when Babur swept through the north of Hindustan early in the sixteenth century, vandalizing Hindu temples, stealing sacred objects, building mosques over temple ruins. For two hundred years, Hindus lived at the mercy of the marauders, until the Maharashtrian hero Shivaji rose up against the aliens and drove them back, restoring the Hindu kingdom. His success, however, was all too brief. Soon the British East India Company and then the British themselves took up where Babur and his progeny had left off, imposing a tyranny upon Hindustan and her people. In short, despite the flawless excellence of indigenous Hindu culture, the men of Hindustan have lived for centuries in a humiliated condition. They can recover pride in themselves only by concerted aggression against alien elements in their midst. If they rise to the occasion, they can restore the original bliss, the time when Rama ruled the world with his blessings.

What is wrong with this picture? Well, for a start, the people who spoke Sanskrit almost certainly migrated into the subcontinent from outside, finding indigenous people there, probably the ancestors of the Dravidian peoples of South India. Hindus are no more indigenous than Muslims. Second, it leaves out problems in Hindu society: the problem of caste, which both Gandhi and Tagore took to be the central social problem facing India, and obvious problems of class and gender inequality. (When historians point to evidence of these things, the Hindu right call them Marxists, as if that by itself invalidated the arguments.) Third, it leaves out the tremendous regional differences within Hinduism, and the hostilities and aggressions sometimes associated with those. Fourth, it omits the evidence of peaceful co-existence and syncretism between Hindus and Muslims over a good deal of the Moghul Empire, including Akbar's well-known policies of religious pluralism.

In the Hindu right's version of history, a persistent theme is that of humiliated masculinity. According to the received story, Hindus have been subordinate for centuries, and their masculinity insulted, in part because they have not been aggressive and violent enough. Even while the violence of the conquerors is decried, Hindu males are encouraged to emulate that aggressive and warlike demeanor.

Rabindranath Tagore, deeply perceptive here as always, represents his Hindu nationalist anti-hero, in *The Home and the World*, as wishing he could seize the woman he desires by force, but finds himself unable to do so. He blames this inability on his Hindu heritage, and wishes for a different nature. He says that there are two different sorts of music, the sound of the Hindu flute and the sound of the British military band. He wishes that he could hear in his blood the sound of the military band, rather than that disturbingly non-aggressive flute.

The two leading ideologues of the Hindu right, who in different ways responded to this call for a warlike Hindu mascuilinity, are V. D. Savarkar, a freedom fighter who spent years in a British prison in the Andaman Islands, and who may have been a co-conspirator in the assassination of Mahatma Gandhi, and M. S. Golwalkar, a gurulike figure who was not involved in the independence struggle and who quietly, behind the scenes, built up the organization called the RSS (National Corps of Volunteers), which is now the leading social organization of the Hindu right. Savarkar's *Hindutva: Who is a Hindu?* (1923) undertakes to define the essence of Hinduness for the new nation; his definition is exclusionary, emphasizing cultural homogeneity and the need to use force to ensure the supremacy of Hindus. For reasons of space, however, I shall focus on Golwalkar's *We: Or, Our Nationhood Defined* (1939). Some of the remarks I am about to quote are embarrassing to Hindus today, and members of the Hindu right hasten to assure one that Golwalkar knew nothing about the Holocaust and withdrew the offending statements in editions of the book published after the war. But 1939, the year of the work's initial publication, was still later than the Nuremberg laws (1935) and Kristallnacht (1938); moreover, my own copy of the fourth edition, published in 1947, still contains the statements, as quoted here.

Writing during the independence struggle, Golwalkar sees his task as describing the unity of the new nation. He announces that most Indians' ideas about nationhood are mistaken:

> They are not in conformity with those of the Western Political Scientists . . . It is but proper, therefore, at this stage to understand what the Western Scholars state as the Universal Nation-idea and correct ourselves. With this end in view, we shall now proceed with stating and analyzing the World's accepted Nation-concept. (1947: 21)

Notice the un-selfconscious deference to European scholarship as what "the World" thinks.

Golwalkar now turns to English dictionaries and to British and German political science. The five elements that he finds repeated as hallmarks of national unity are: geography, race, religion, culture and language. Golwalkar examines each of these in turn and then analyzes several nations to see the extent to which they embody the desired unities. Germany impresses him especially for the way in which it has managed to bring "under one sway the whole of the territory" that was originally

held by the *Germani* but was parceled out under different regimes (42). Turning to race, he observes:

> German race pride has now become the topic of the day. To keep up the purity of the Race and its culture, Germany shocked the world by her purging the country of the semitic Races—the Jews. Race pride at its highest has been manifested here. Germany has also shown how well nigh impossible it is for Races and cultures, having differences going to the root, to be assimilated into one united whole, a good lesson for us in Hindusthan to learn and profit by. (43)

In the end, Golwalkar's vision of national unity is not exactly that of Nazi Germany. He is not very concerned with purity of blood, and far more concerned with a group's desire to merge into the dominant whole. Groups who fall outside the five-fold definition of nationhood, he concludes, can "have no place in the national life . . . unless they abandon their differences, and completely merge themselves in the National Race. So long, however, as they maintain their racial, religious and culture differences, they cannot but be only foreigners, who may be either friendly or inimical to the Nation" (53). Unlike Hitler, Golwalkar would probably be happy with the conduct of the many German Jews who converted to Christianity and assimilated their lifestyle to the dominant German lifestyle. Nonetheless, he is firmly against the civic equality of any people who retain their religious and ethnic distinctiveness, refusing to merge into the dominant Hindu whole. He speaks approvingly of the idea that people who refuse to assimilate should lose their civil rights, living "at the sufferance of the Nation and deserving of no special protection, far less any privilege or rights" (55). Here is the way in which Golwalkar applies his ruminations about the "old nations" of Europe to the case of India:

> There are only two courses open to the foreign elements: either to merge themselves in the national race and adopt its culture, or to live at the sweet will of the national race. That is the only logical and correct solution. That alone keeps the national life healthy and undisturbed. That alone keeps the Nation safe from the danger of a cancer developing into its body politics [*sic*] of the creation of a state within the state. From this standpoint, sanctioned by the experiences of shrewd old nations, the non-Hindu peoples in Hindusthan must either adopt the Hindu culture and language, must learn to respect and hold in reverence Hindu religion, must entertain no idea but those of glorification of the Hindu race and culture i.e. they must not only give up their attitude of intolerance and ungratefulness towards this land and its agelong traditions but must also cultivate the positive attitude of love and devotion instead—in one word they must cease to be foreigners, or may stay in the country, wholly subordinated to the Hindu nation, claiming nothing, deserving no privileges, far less any preferential treatment—not even citizen's rights. There is, at least should be, no other course for them to adopt. (55–6)

At this time, the RSS was a cultural movement without a distinct political arm. It disavowed any connection with the more political Hindu Mahasabha, which later did propose that Muslims and Christians should lose all civil rights in the new

nation. Clearly, however, this was Golwalkar's program, and we should not doubt that such aims lie underneath the modern political arm of the RSS, the political party known as the BJP, or Bharatiya Janata Party (National People's Party).

At the time of independence, such ideas of Hindu supremacy did not prevail. Nehru and Gandhi insisted not only on equal rights for all citizens but on the most stringent protection for religious freedom of expression in the new constitution. Gandhi always pointedly included Muslims at the very heart of his movement, and a devout Muslim, Maulana Azad, was not only one of his and Nehru's most trusted advisors; he was the person to whom Gandhi turned to accept food when he broke his fast unto death, a very pointed assault on sectarian ideas of purity and pollution. Such ideas never went uncontested.

On 30 January 1948, Mahatma Gandhi was shot at point-blank range by Nathuram Godse, a member of the Hindu Mahasabha and former member of the RSS. Godse, who edited a newspaper called *Hindu Rashtra* (Hindu Nation), had left the RSS because it seemed to him not political enough; the Mahasabha, a political party, was more congenial. As was shown by a letter written by Godse to Savarkar in 1938 and submitted to the trial court, Godse had long had a close relationship with Savarkar, whom he revered. "Since the time you were released from your internment at Ratnagiri," he wrote, "a divine fire has kindled in the minds of those groups who profess that Hindustan is for the Hindus" (quoted in Ramachandran 2004). He speaks of using the Hindu Mahasabha (of which Savarkar was then President) to build a national volunteer army, drawing on the resources of the RSS, where Godse was then a leading local organizer.

At his sentencing on 8 November 1949, Godse read a long (book-length) statement of self-explanation, justifying his assassination for posterity (see Elst 2001). Although the statement was not permitted publication at the time, it gradually leaked out into the public realm. Translations into Indian languages began appearing, and in 1977 the English original was published by Godse's brother Gopal under the polite title, *May it Please your Honour*. A new edition, with a long epilogue by Gopal, was published in 1993 under the more accurate title *Why I Assassinated Mahatma Gandhi* (see Elst 2001: 5–6). Today the statement is also widely available on Hindu right websites where Godse is revered as a hero, and, on one website entirely devoted to his career (www.nathuramgodse.com), as "The True Patriot and the True Indian".[2] (This website also contains the text of a recent Marathi-language play glorifying Godse that has been banned in India.)

Godse's self-justification, like the historical accounts of both Savarkar and Golwalkar, sees recent events against the backdrop of centuries of "Muslim tyranny" in India, punctuated by the heroic resistance of Shivaji. Like Savarkar, Godse describes his goal as that of creating a strong, proud India that can throw off

[2] See <www.nathuramgodse.com>. Quotations from Godse's statement in what follows are taken from this website.

the centuries of domination. Godse is appalled by Gandhi's rejection of the warlike heroes of classical Hindu epics: "It is my firm belief that in dubbing Rama, Krishna and Arjuna as guilty of violence, the Mahatma betrayed a total ignorance of the springs of human action." Indeed, he argues, it is Gandhi who is the more guilty of violence, since he exposes Indians to subordination and humiliation: "He was, paradoxical as it may appear, a violent pacifist who brought untold calamities on the country in the name of truth and non-violence, while Shivaji [and other resistance fighters] will remain enshrined in the hearts of their countrymen for ever for the freedom they brought to them." (So deep was Godse's objection to non-violence that he earlier refused the offer to commute his sentence to life imprisonment, saying, "Please, see to it that mercy is not imposed on me. I want to show that through me, Gandiji's non-violence is being hanged.")

Godse's second major objection to Gandhi is to his "pro-Muslim policy", which he sees in many aspects of Gandhi's politics, for example his support for Urdu alongside Hindi as national languages,[3] and his willingness to placate Jinnah and the Muslim League. Gandhi, he argues, has betrayed his role as father of the Indian nation and has become the father of Pakistan.

Godse tells us that he gradually came to the conclusion that Gandhi's disastrous policies could only be brought to an end by ending Gandhi's life. Such was Gandhi's personal charisma that so long as he lived, the Congress Party would have to "be content with playing second fiddle to all his eccentricity, whimsicality, metaphysics and primitive vision". Gandhi's "childish insanities and obstinacies, coupled with a most severe austerity of life, ceaseless work and lofty character made Gandhi formidable and irresistible". So he planned in secret, he says, telling nobody about his plans, and fired the fatal shots.

Toward the end of Godse's statement appears a passage that heads the Hindu right website devoted to his memory:

> If devotion to one's country amounts to a sin, I admit I have committed that sin. If it is meritorious, I humbly claim the merit thereof. I fully and confidently believe that if there be any other court of justice beyond the one founded by the mortals, my act will not be taken as unjust. If after the death there be no such place to reach or to go, there is nothing to be said. I have resorted to the action I did purely for the benefit of the humanity. I do say that my shots were fired at the person whose policy and action had brought rack and ruin and destruction to lakhs [hundreds of thousands] of Hindus.

Nehru believed that the murder of Gandhi was part of a "fairly widespread conspiracy" on the part of the Hindu right to seize power (a letter of 1948, quoted in Jaffrelot 1996: 87); he saw the situation as analogous to that in Europe on the eve

[3] Hindi and Urdu are not very different as languages; at most they are slightly different dialects. The major difference between them is the script in which they are written: Persian script in the case of Urdu; Devanagari (the Sanskrit script) in the case of Hindi. Thus it is odd to apply ideas of linguistic nationalism to this question.

of the fascist takeovers. And he believed that the RSS was the power behind this conspiracy. In December 1947, he had already written to the provincial governors:

> We have a great deal of evidence to show that the RSS is an organization which is in the nature of a private army and which is definitely proceeding on the strictest Nazi lines, even following the technique of organization . . . I have some knowledge of the way the Nazi movement developed in Germany. It attracted by its superficial trappings and strict discipline considerable numbers of lower middle class young men and women who are normally not too intelligent and for whom life appears to offer little to attract them.
>
> (quoted in Jaffrelot 1996: 87)

For reasons of space, we must now fast-forward to recent years. Although illegal for a time, the RSS eventually re-emerged, and quietly went to work building a vast social network, consisting largely of groups for young boys, called "branches" or *shakas*, which, through clever use of games and songs, indoctrinate the young into the confrontational and Hindu-supremacist ideology of the organization. The idea of total obedience and the abnegation of the critical faculties is at the core of this solidaristic movement. Each day, as members raise the saffron flag of the warlike hero Shivaji, which the movement prefers to the tricolor flag of the Indian nation, with its Buddhist wheel of law reminding citizens of the emperor Ashoka's devotion to religious toleration, they recite the following pledge: "I take the oath that I will always protect the purity of Hindu religion, and the purity of Hindu culture, for the supreme progress of the Hindu nation. I have become a component of the RSS. I will do the work of the RSS with utmost sincerity and unselfishness and with all my body, soul, and resources. And I will keep this vow for as long as I live. Victory to Mother India." The organization also makes clever use of modern media: a nationally televised serial version of the classical epic *Ramayana* in the late 1980s fascinated viewers all over India with its concocted tale of a unitary Hinduism dedicated to the single-minded worship of the god Rama and to his birthplace at Ayodhya in North India. As a result of the propaganda stirred up in this and other ways, in 1992 Hindu mobs, with the evident connivance of the modern political wing of the RSS, the BJP, destroyed a mosque in the city of Ayodhya which they say covers the remains of a Hindu temple marking Rama's birthplace.

Meanwhile, politically, the BJP began to gather strength in the late 1980s, drawing on widespread public dissatisfaction with the economic policies of the post-Nehru Congress Party (although it was actually Congress, under Rajiv Gandhi, who began economic reforms), and playing, always, the cards of hatred and fear. It was during their rise to power, and shortly before they were able to govern in a coalition government (which prevented them from carrying out all of their goals), that the destruction of the Ayodyha mosque took place. The violence in Gujarat was the culmination of a series of increasingly angry pilgrimages to the Ayodyha site, where the Hindu right has attempted to construct a Hindu temple over the ruins, but has

been frustrated by the courts. Although, as noted earlier, the elections of 2004 gave a negative verdict on the BJP government, they remain the major opposition party and control state governments in some key states, including Gujarat.

III. Lessons of Gujarat

For several years I have studied the Gujarat violence, its basis and its aftermath, looking for its implications for the ways in which we should view religious violence around the world. One obvious conclusion to draw is that each case must be studied on its own merits, with close attention to specific historical and regional factors. The idea that all conflicts are explained by a simple hypothesis of the "clash of civilizations" proves utterly inadequate in the face of Gujarat, where European ideas, borrowed to address a perceived humiliation, have been used to create an ideology that has ultimately led to a great deal of violence against peaceful Muslims, and to the threat of more violence to come. Indeed, the clash-of-civilizations thesis is the best friend of the perpetrators, because it shields them and their ideology from scrutiny. Repeatedly in interviews with leading members of this group, I was informed that no doubt, as an American, I was already on their side, knowing that Muslims cause trouble wherever they are, and that there is nothing for it but to be prepared to take reactive, and even pre-emptive, measures against them. What we see in Gujarat is not that simplistic, comforting thesis, it is something more disturbing: the fact that in a thriving democracy many individuals are unable to live with others who are different, on terms of mutual respect and amity. They seek total domination as the only road to security and pride. This is a phenomenon that is well known in democracies around the world, and it has nothing to do with an alleged Muslim monolith, and, really, very little to do with religion as such. (I've noted that the Hindu right are bent on recasting Hinduism into a newly violent and monolithic form.)

This case, then, informs us that we must look within, asking whether similar forces are at work in our own society, and, if so, what we might do to counteract them. Beyond that general insight, my study of the riots has suggested four very specific lessons.

III.1 The Rule of Law

One of the most appalling aspects of the Gujarat pogrom was the complicity of officers of the law. Police sat on their hands, and indeed were ordered to do so,

on penalty of demotion or transfer. The highest officials of state government, prominently including Chief Minister Narendra Modi, egged on the killing. The national government gave aid and comfort to the state government. The worst aspect of the unequal status of Muslims in Gujarat was this inequality before the law: Muslims at present are not equal citizens.

The institutions of government broke down at the local level, and to some extent at the national level. However, the institutional and legal structure of the Indian democracy ultimately proved robust, playing a key role in securing justice for the victims. The Supreme Court and the National Electoral Commission played very constructive roles in postponing new elections while Muslims were encouraged to return home, and in ordering changes of venue in key trials arising out of the violence. Above all, there were free national elections in 2004, and these elections, in which the participation of poor rural voters was decisive, delivered a strongly negative verdict on the policies of fear and hate, as well as the BJP's economic policies. The current government, headed by Manmohan Singh, India's first minority Prime Minister, has announced a firm commitment to end sectarian violence and has done a great deal to focus attention on the unequal economic and political situation of Muslims in the nation, as well as appointing Muslims to key offices. On balance, then, the pluralistic democracy envisaged by Gandhi and Nehru seems to be winning, in part because the framers bequeathed to India a wise institutional and constitutional structure and traditions of commitment to the key political values this structure embodies.

It should be mentioned that one of the key aspects of the founders' commitments, which so far has survived the challenge from the Hindu right, is the general conception of the nation and its unity as a unity around political ideals and values, particularly the value of equal entitlement, rather than around ethnic or religious or linguistic identity. India, like the United States, but unlike most of the nations of Europe, has rejected such exclusionary ways of characterizing the nation, adopting, in the constitution, in public ceremonies, and in key public symbols, the political conception of its unity. India's national anthem, whose words and music were composed by Rabindranath Tagore, is a paean to pluralism: it mentions the diverse ethno-geographical origins of Indians, and praises the fact that all alike show reverence to the moral law. This public commitment to pluralism is continually contested by the Hindu right, who prefer another anthem, warlike and exclusionary. So far, however, the founders' strategy has held firm. Political structure is not everything, but it can supply a great deal in times of stress.

III.2 The Press and the Role of Intellectuals

One of the heartening aspects in Gujarat was the performance of the national media and of the community of intellectuals. Both print media and television kept

up unceasing pressure to document and investigate the riots, and the role of key government officials was documented beyond doubt. At the same time, because the local police were not doing their job, many scholars, lawyers and NGO leaders converged on Gujarat to take down the testimony of witnesses, help them file complaints, and prepare a public record that would stand up in court. The intellectual community has easier access to the national press in India than in the US, in part thanks to the somewhat greater financial independence of the national media, and these intellectuals seized the opportunity, producing a wonderful outpouring of trenchant and high-quality analysis. The only reason I felt the need to write further about these events myself is that these analyses have by and large not reached the US audience. We can see here a documentation of something long ago observed by Amartya Sen in the context of famines: the crucial role of a free press in supporting democratic institutions. And we can study here what freedom really means: I would argue that it requires a certain absence of top-down corporate control and easy access to the major media for intellectual voices from a wide range of backgrounds. We in the US should take note.

Sen has recently returned to the topic of public debate in analyzing Indian politics, and I utterly agree with the two major contentions of *The Argumentative Indian* (Sen 2005): first, that India has a long and glorious tradition of public debate, and second, that this tradition is in jeopardy, given the ideological and anti-rational tactics of the Hindu right. Here I would only add the fact that market forces are usually not kind to investigative journalism, and that all nations need to beware lest the crucial contribution of a free press to democracy be weakened beyond repair (see Baker 2007).

III.3 Education: The Importance of Critical Thinking and Imagination

So far I have mentioned factors that have helped the Indian democracy survive the threat of quasi-fascist takeover. Now we move to warning signs for the future, areas in which the democracy is currently weak and vulnerable. The publicly funded schools of the state of Gujarat are famous for their complete lack of critical thinking, their exclusive emphasis on rote learning and the uncritical learning of marketable skills, and for the elements of fascist propaganda that easily creep in when critical thinking is not cultivated. It is well known that Hitler is presented as a hero in history textbooks in this state, and nationwide public protest has not yet led to any change in this regard. To some extent the rest of the nation is better off than Gujarat: national-level textbooks have been rewritten to take out the false ideological view of history beloved of the Hindu right and to substitute a much more nuanced view of history. Nonetheless, the emphasis on rote learning and on regurgitation

for national examinations is distressing everywhere, and things are only becoming worse with the immense pressure to produce economically productive graduates. The educational culture of India used to contain progressive voices, such as the great Tagore, who emphasized that all the skills in the world were useless, even baneful, if not wielded by a cultivated imagination and refined critical faculties. Currently, these voices have been silenced by the sheer demand for profitability in the global market. Parents want their children to learn marketable skills, and their great pride is the admission of a child to the Indian Institutes of Technology and Management. They hold the humanities and the arts in contempt. I fear for democracy down the road, when it is run, as it increasingly will be, by docile engineers in the Gujarat mold, unable to criticize the propaganda of politicians and unable to imagine the pain of another human being.

This is no humorous topic, but it can be illustrated by an odd story from my own experience investigating the Gujarati community in the United States, where fully 40% of Indian-Americans hail from that state. A large proportion of Gujarati Hindus belong to the Swaminarayan sect of Hinduism, which at present is distinctive for its emphasis on uncritical obedience to the utterances of the current head of the sect, whose title is Pramukh Swami (see Nussbaum 2007a: ch. 9). On a visit to the elaborate multi-million-dollar Swaminarayan temple in Bartlett, Illinois, I was given a tour by a young man recently arrived from Gujarat, who delighted in telling me the simplistic Hindu right story of India's history, and who emphatically told me that whenever Pramukh Swami speaks, one is to regard it as the direct voice of God and obey without question. At this point, with a beatific smile, this young man pointed up to the elaborate marble ceiling of the temple and asked me, "Do you know why this ceiling glows the way it does?" I said I didn't know, and I confidently expected an explanation invoking the spiritual powers of Pramukh Swami. My guide smiled even more broadly. "Fiber-optic cables", he told me. "We are the first ones to put this technology into a temple." Here you see what can easily wreck democracy: a combination of technological sophistication with utter docility. I fear that many democracies around the world, including our own, are going down this road, through a lack of emphasis on the humanities and arts and an unbalanced emphasis on profitable skills (see also Nussbaum 2007b).

III.4 The Creation of A Liberal Public Culture

How did fascism take such a hold in India? Hindu traditions emphasize tolerance and pluralism, and daily life in India, as in New York, tends to emphasize the ferment and vigor of difference, as people from so many different ethnic, linguistic, and regional backgrounds encounter one another. But the traditions contain a wound, a locus of vulnerability, and I would locate this wound in the area of

humiliated masculinity. For centuries, some Hindu males think, they have been subordinated, laughed at, treated as weak by a sequence of conquerors (Nussbaum 2007a: ch. 6). The fact that the British really did despise Hinduism as what Winston Churchill called a "beastly religion" surely made matters worse, and Hindus came to identify the sexual playfulness and sensuousness of their traditions, scorned by the masters of the Raj, with their own weakness and subjection. So a repudiation of the sensuous and the cultivation of the masculinity typified by Tagore's image of the British military band came to seem the best way out of subjection. One reason why the RSS attracts such a following is this widespread sense of masculine failure, a key aspect of the rhetoric of both Savarkar and Golwalkar, and of RSS *shakas* in every part of India today.

At the same time, the RSS filled a void, organizing at the grass-roots level with great discipline and selflessness. The RSS is not just about fascist organization: it also provides needed social services, and it provides fun, luring boys in with the promise of a group life that has both more solidarity and more imagination than the tedious world of government schools. Golwalkar said that if he saw a beautiful peacock in his garden and wanted the peacock to become his pet, he would feed it little bits of opium until it became addicted, and that way it would come to his garden every day (Golwalkar 1966: 348). He said that this was how the *shakas* work: by the lure of fun and games, they make boys obedient members of the organization.

So what is needed is some counter-force which would supply a public culture of pluralism with equally efficient grass-roots organization, and a public culture of masculinity that would contend against the appeal of the warlike and rapacious masculinity purveyed by the Hindu right. The "clash within" is not so much a clash between two groups in a nation who are different from birth; it is, at bottom, a clash within each person, in which the ability to live with others on terms of mutual respect and equality contends fretfully against anxiety and the sense of being humiliated. Gandhi understood this. During his lifetime, his powerful movement did purvey a counter-image to the images of domineering masculinity. He taught his followers that life's real struggle was a struggle within the self, against one's own need to dominate and one's fear of being vulnerable. He deliberately focused attention on sexuality as an arena in which domination plays itself out with pernicious effect, and he deliberately cultivated an androgynous maternal persona. I think that in some respects he went off the rails, in his suggestion that sexual relations are inherently scenes of domination and in his recommendation of asceticism as the only route to non-domination. Nonetheless, he saw the problem at its root, and he proposed a public culture that, while he lived, was sufficient to address it. His followers understood that being a real man does not mean emulating British aggressiveness and learning to bash others. It meant having the courage not to bash, to stand up to aggression with nothing but one's naked human dignity around one. In the process, he won the respect of the entire world for India's men and their

traditions (conceived as he conceived them). In a quite different way, Rabindranath Tagore also created a counter-image of the Indian self, an image that was more sensuous, more joyful than that of Gandhi, but equally bent on renouncing the domination that Tagore saw as inherent in European traditions.

Since Gandhi, however, this part of the pluralist program has languished. Much though he loved and admired both Gandhi and Tagore, Nehru had contempt for religion, and out of his contempt he neglected the cultivation of that which the radical religions of both men had supplied: images of who we are as citizens, symbolic connections to the roots of human vulnerability and openness, and the creation of a grass-roots public culture around these symbols. Nehru was a great institution-builder, but in thinking about the public culture of the new nation his focus was always on economic issues, never on cultural issues. Because he firmly expected that raising the economic level of the poor would cause them to lose the need for religion and in general for emotional sources of nourishment, he saw no need to provide a counterforce to the powerful emotional propaganda of the Hindu right. Today's young people in India, therefore, tend to think of religion, and symbolic culture-creation in general, as forces that are in their very nature fascist and reactionary, because that is what they have seen in their experience. When one tells them the story of the US civil rights movement, and the role of both liberal religion and powerful pluralist rhetoric in forging an anti-racist civic culture in that movement, they are quite surprised. Meanwhile, the RSS, which understands human psychology rather well, goes to work unopposed in every state and region, skillfully plucking the strings of hate and fear. By now pluralists generally realize that a mistake was made in leaving grass-roots organization to the right, but it is very difficult to jump-start a public pluralist movement. The salient exception has been, for some years, the women's movement, which has built at the grass-roots very skillfully, with the regional knowledge, the mixture of economic and cultural incentives, and the respect for the arts and the imagination that the creation of such a movement requires.

Here I believe that I do have some serious differences with Amartya Sen's analysis of religious violence (Sen 2006). Sen suggests that religious violence is based on an intellectual mistake: people are bamboozled into thinking that there is a single identity that they simply must make their core self-definition, and they are then convinced that this identity "makes extensive demands... (sometimes of a most disagreeable kind)" (p. xii). Instead, they should wake up and realize that they have lots of choices, and that leading a human life centrally involves "the responsibilities of choice and reasoning" (p. xii). What is needed in order to set things on a better course is simply to remind people of their freedom and show them that there is no inevitability about the "destiny" of being associated with a particular ethnic or religious group and its leaders' demands.

Now of course I am all for public debate, and I also agree heartily with Sen that people have many choices about how to identify themselves. I think, however,

that his diagnosis remains on the surface, and doesn't come to grips with what makes otherwise reasonable and intelligent people go in for violent behavior. His unwillingness to confront the murkier depths of human psychology makes his paean to reasoning seem, actually, a bit condescending: the world is in trouble only because people haven't realized some pretty obvious facts.

People are not stupid, however, and bad behavior is not based on an easily corrected error. We need to try to understand what it is about the experience of humiliation and subordination that leads to retaliatory aggression. So we need to talk about the weakness that all human beings share, their shame before that weakness, their disgust at the parts of the human body that are the keenest reminders of weakness (Nussbaum 2001: ch. 4; 2004). And then we need to talk about masculinity and how constructed images of masculinity express, in some cases, disgust, shame and retaliatory aggression—and, in other cases, a sense of dignified acceptance of the limited character of human existence. Gandhi's genius was to understand where hatred resides and to see that one cannot dislodge it without proposing new images of what a real man is, what a real human being is. His contention that we cannot undo hatred without undertaking a total reform of emotions connected to masculine pride seems to me utterly correct (Nussbaum 2007a: ch. 6). Tagore had similar insights, perhaps with a healthier and more optimistic understanding of the potential of human sexuality. He too thought that images of the real man needed to be remade if human beings were ever going to realize the humanity of which they were capable. Such proposals of internal critique and reform find no place in Sen's impatiently intellectualistic reading of violence.

Sen's diagnosis and mine lead in the same direction to some extent, in that both of us favor a great emphasis on argument and critical thinking, both in schools and in the general public culture. My understanding of violence, however, requires far more: a large role for the cultivation of sympathy in schools (as Tagore insistently proposed), together with a large role for the humanities and the arts; and then—what must surround schools if their work is not to be futile—a civic culture with symbolic and artistic elements linking manliness to restraint and compassion. This program is not at all easy to achieve; to that extent my own "take" on the future of democracy in the world remains less confident than Sen's.

It is comforting for Americans to talk about a clash of civilizations. This thesis tells us that evil is outside, distant, other, and that we are perfectly all right as we are. All we need do is to remain ourselves and fight the good fight. But the case of Gujarat shows us that the world is very different from the world as depicted in that comforting fable. The forces that assail democracy are internal to many if not most democratic nations, and they are not foreign: they are our own ideas and voices, meaning the voices of aggressive European nationalism, refracted back against the original aggressor with the extra bile of resentment born of a long experience of domination and humiliation. The implication of this idea is that all nations, Western and non-Western, need to examine themselves with the most

fearless exercise of the critical capacities, looking for the roots of domination within and devising effective institutional and educational counter-measures. At a deeper level, the case of Gujarat shows us what Gandhi and Tagore in their different ways knew very well: that the real root of domination lies deep in the human personality, in the narcissistic desire to dominate others and to efface the inconvenient challenge posed by the other; in wounded masculinity that cannot rest until it has destroyed the source of its own perceived wound. It would be so convenient if Americans were pure and free from flaw, but we can now see that very fantasy of purity for what it is: as yet another form that the resourceful narcissism of the human personality takes on the way to bad behavior.

Looking at the clash in my way, as an internal clash, we will naturally focus on four strategies for the preservation and enhancement of democracy around the world: first, on getting institutional structures that can remain firm against fascist challenges; second, on bolstering independence, including the economic independence of the press and the free speech of intellectuals; third, on creating educational institutions that teach the skills of critical thinking and imagining that are so crucial for the health of democracy; and finally—what Martin Luther King Jr. learned from Gandhi—on creating a public culture of non-domination and equality that can inspire fearful human beings, for all of us are fearful, with the idea that comfort is to be found in mutual aid and reciprocity, rather than quick and dirty victory over an enemy onto whom we have all too conveniently projected our own fears.

References

Baker, R. (2007), "Goodbye to Newspapers?", *New York Review of Books*, 54(13).

Elst, K. (2001), *Gandhi and Godse: A Review and a Critique* (Delhi: Voice of India).

Golwalkar, M. S. (1947), *We: Or, Our Nationhood Defined*, 4th edn (Nagpur: P. V. Betwalkar; 1st edn 1939).

—— (1966), *Bunch of Thoughts* (Bangalore: Jagarana Prakashan).

Huntington, S. P. (1996), *The Clash of Civilizations and the Remaking of World Order* (New York: Simon and Schuster).

Jaffrelot, C. (1996), *The Hindu Nationalist Movement in India*, tr. of *Les nationalistes hindous* (New York: Columbia University Press).

Kennan, G. F. (1990), "Comments on the National Security Problem", in G. D. Harlow and G. C. Maerz (eds), *Measures Short of War: The George F. Kennan Lectures at the National War College* (Washington, DC: National Defense University Press).

Nussbaum, M. C. (2001), *Upheavals of Thought: The Intelligence of Emotions* (New York: Cambridge University Press).

—— (2004), *Hiding from Humanity: Disgust, Shame, and the Law* (Princeton, NJ: Princeton University Press).

—— (2007a), *The Clash Within: Democracy, Religious Violence, and India's Future* (Cambridge, Mass.: Harvard University Press).
—— (2007b), "The Robot Corporation", *Outlook*, 20 August: 112–13.
Ramachandran, R. (2004), "The Mastermind", *Outlook*, 6 September: available at <http://communalism.blogspot.com/2004/08/india-mastermind-behind-gandhis.html>.
Savarkar, V. D. (1969), *Hindutva: Who is a Hindu?*, 5th edn (Bombay: S. S. Savarkar; 1st publ. 1923).
Sen, A. K. (2005), *The Argumentative Indian* (London: Allen Lane).
—— (2006), *Identity and Violence: The Illusion of Destiny* (New York: W. W. Norton).

CHAPTER 28

ENGAGING WITH IMPOSSIBILITIES AND POSSIBILITIES

ELINOR OSTROM

In the lecture that Amartya Sen delivered on the occasion of receiving the Alfred Nobel Memorial Prize in Economic Sciences (Sen 1999), he devoted considerable attention to Kenneth Arrow's (1951, 1963) "impossibility theorem". Arrow had elegantly demonstrated that short of a dictatorship, social choice procedures could not always generate unambiguous and consistent outcomes for a group of rational individuals who had their own preference rankings over collective outcomes that simultaneously met a set of reasonable conditions. Arrow used the definition of rationality accepted by most economists in the mid-twentieth century. He assumed that individuals making a social choice had well-defined, complete and transitive preferences regarding the likely net personal returns given the set of alternatives under consideration.

Sen argued against the grim predictions that social choice theory had been destroyed. Instead, he posited that Arrow's theorem should generate engagement rather than resignation. The theoretical results deserve serious study, Sen argued, given their sweeping reach, even though he agreed that democratic decisions do not always generate good results and that they can at times lead to incongruities:

> To the extent that this is a feature of the real world, its existence and reach are matters for objective recognition. Inconsistencies arise more readily in some situations than in others, and it is possible to identify the situational differences and to characterize the processes through which consensual and compatible decisions can emerge. (Sen 1999: 365)

Sen has followed his own advice creatively and seriously and effectively engaged Arrow's "impossibility theorem" over time (see e.g. Sen 1977b, 1986, 1993, 2002). Instead of despair, he has examined both the underlying assumptions of individual rationality (1977a) and the empirical conditions under which diverse voting rules make a major difference in outcomes.[1]

In this chapter, I wish to honor Amartya Sen by illustrating how his advice to engage with impossibility results, rather than dismissing or accepting them without question, has been an important inspiration related to another widely acclaimed impossibility result—that of Garrett Hardin (1968) in his influential article in *Science* on "The Tragedy of the Commons". While Hardin was an ecologist rather than an economist, his assumptions regarding individual preferences and behavior are closely aligned with those of many economists, focusing as they do on immediate material returns to self. His logic was broadly similar to that of the distinguished economist H. Scott Gordon, who had earlier argued that "The fish in the sea are valueless to the fisherman, because there is no assurance that they will be there for him tomorrow if they are left behind today" (1954: 124).

Hardin envisioned a pasture open to all, in which each herder received an individual benefit from adding animals to graze on the pasture and suffered only delayed costs (with his fellow herders) from overgrazing. He assumed that there were no property rights to the land and no specified rights related to grazing on the land.[2] Hardin concluded:

> Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit—in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. (Hardin 1968: 1244)

Hardin's conclusion was immediately accepted by many scientists from multiple disciplines. His article is still required reading in most environmental science programs and frequently assigned multiple times during an undergraduate program. It was the article most frequently cited as having the greatest career impact in a recent survey of biologists undertaken by Barrett and Mabry (2002). For many economists, Hardin's theory was accepted as an extreme example of the theory developed by Mancur Olson (1965) on the logic of collective *in*action (if I may be pardoned for a play on words).

Policy-makers also tended to accept Hardin's and Olson's results and thought their conclusions made it obligatory to take positive action to impose rules on the users of pastures, forests, fisheries, water systems, and other common-pool

[1] See, for example, Sen (1981, 1984), where he shows that no famine has ever occurred in an independent country except when the country is run by a dictator, an alienated set of rules, or a single party.

[2] Hardin (1998) did clarify thirty years later that he had meant an "open-access commons", but many readers do not see any significant difference between open-access regimes and a common-pool resource with a common-property regime.

resources (CPRs) in their domain. Government officials did not examine whether the users of these resources had developed rules of their own over time, because it was simply assumed that resource users were trapped and helpless. National governments around the world declared that government ownership was the *only* way to save resources from destruction. Forests in India, Thailand and many countries in Africa were nationalized during the 1970s and 1980s after the immediate acceptance of the notion that this was the *only* way to avoid massive deforestation (Feeny 1988; Arnold and Stewart 1991; Shepard 1992; Thomson *et al.* 1992). In many instances, this, and the related conversion of inshore fisheries policies, had the effect of overruling locally developed institutions on the ground and converting these resources to "open access", given that there was little administrative follow-up and that local users were told they had no rights (Cordell and McKean 1992; Higgs 1996).

I. Engaging Hardin

Before Hardin published his impossibility results, as my dissertation research I had conducted an in-depth study of how groundwater producers had grappled with serious problems resulting from their own overuse of groundwater underlying the Los Angeles metropolitan area in the 1960s (Ostrom 1965). The 500-plus producers were located in parts of a dozen general-purpose local governments but none of these governmental jurisdictions had boundaries that were similar to those of the underlying groundwater basins. The producers were fortunate, however, to be in a state that encouraged diverse forms of self-governance, to have access to expert technical information from the US Geological Survey (USGS) and the California State Department of Natural Resources, and to be able to utilize the California courts in a series of cases utilizing equity jurisprudence.[3]

Watching these efforts, I gained considerable respect for the capability of *public* entrepreneurs in creating new institutional arrangements to cope relatively effectively with severe overdrafts and the threat of losing the high value of a groundwater basin to salt-water intrusion from the Pacific Ocean. The producers did utilize public facilities in the form of equity jurisprudence in the California courts and the capacity to establish special districts within the California legal framework (Blomquist 1989, 1992; Blomquist and Ostrom 2008). The producers were not forced by an external authority, however, to take these actions.

[3] The producers contributed a required portion of the costs to the studies undertaken by the USGS as well as to the cost of the equity jurisprudence proceedings in the courts. Thus, as long as they were willing to contribute a portion of the costs, they were able to obtain accurate information compiled by third-party experts.

During the 1970s and 1980s, colleagues at the Workshop in Political Theory and Policy Analysis at Indiana University developed a framework for studying institutions called Institutional Analysis and Development (IAD) (Kiser and Ostrom 1982; Oakerson 1992; Ostrom 2005). A key element of that framework distinguished among three levels of rules. Operational rules are the rules that individuals or groups use in daily life and affect real-world phenomena directly. They include the speed limits set on roads of various classes, the hours that a private firm is open for business, the rules that define who can use a resource and the conditions of its use. Collective-choice rules are those that are used in making operational-choice rules. Elections, most decisions made in a legislature, and many court rulings exist at a collective-choice level. Constitutional-choice rules are the level of rules used to make the rules of collective-choice decision-making (Buchanan and Tullock 1962). These three levels are used in all social groups and political organizations, ranging in spatial terms from the very small to the global (see Aoki 2007b for a similar analysis of the multiple levels of rules). This framework opens up the possibility that resource users could change their own rules.

By the early 1980s, the widespread acceptance of the gloomy predictions of both Olson (1965) and Hardin (1968) had led the governments of multiple developing countries to nationalize common-pool resources. Overuse tended to *increase* after nationalization, however. Scholars began to gather evidence that demonstrated how some local users had actually overcome the problems they faced (Netting 1972, 1981; Berkes 1985; McCay and Acheson 1987). It also led to the creation of a National Research Council (NRC) committee to study diverse common-pool resources (NRC 1986). As a member of the NRC committee, I began to recognize the *disparity* between the impossibility result and the repeated evidence that many (but certainly not all) resource users had found ways to change their own institutions. Since then, I have been actively engaged in coming to grips with the impossibility results of Garrett Hardin and the gloomy predictions of Mancur Olson. Let me provide a quick review of this engagement and present a theory of self-organization that is more representative than Hardin's tragedy of the problem faced by resource users.

II. Why Some Users Overcome Common-Pool Resource Dilemmas and Others Do Not

When dealing with a presumption of an *impossibility* of resource users solving their own problems of overuse, finding a very large number of cases where resource users succeed is an important accomplishment. That is exactly what the NRC

committee did. A major conceptual accomplishment of the NRC committee was the emphatic recognition of the long-standing confusion over the names used for resources and for potential institutional arrangements or property regimes. The term "common property resource" was abbreviated as CPR and widely used across the social sciences for resources such as forests, lakes, pastures, fisheries and irrigation systems. But this term confused a resource system that might or might not have a linked property-rights system with a form of institution called "common property" (Ciriacy-Wantrup and Bishop 1975; Ostrom and Ostrom 1977; Bromley 1992). That is, it confused a resource—a common-pool resource—with a property system—a common-property regime. The initials CPR were used for both concepts.

Slowly, over time, general agreement has been reached that common-pool resources have two characteristics:

1. it is very costly to exclude potential beneficiaries from accessing and harvesting the resources, and
2. the amount of resource flows harvested by one user is subtracted from the quantity available to others.

A common-pool resource shares the first characteristic with public goods (the cost of exclusion) and the second characteristic with private goods (subtractability) (Ostrom and Ostrom 1977).

In Hardin's illustrative case of "the tragedy of the commons", the common-pool resource is "open access", i.e. owned by no one. But open access is only one of four general possibilities for the use of a common-pool resource. Such a resource can also be managed under any of three kinds of property-rights regimes:

- government ownership, where a formal government ranging in size from a local city all the way up to national government claims ownership of the resource and the right fully to determine who can or cannot use it and under what circumstances;
- private ownership, where a single individual or private firm has full claims to determine use patterns; and
- community or common-property ownership, where a group of individuals shares rights to ownership.

As Bromley has so clearly stated:

> There is no such thing as a common property *resource*; there are only resources controlled and managed as common property, or as state property, or as private property. Or—and this is where confusion persists in the literature—there are resources over which *no property rights* have been recognized. We call these latter "open-access resources" (*res nullius*, which is Latin for "no one's property"). (Bromley 1992: 4; see also Bromley 1989, 1991)

In addition to conceptual clarification, the NRC (1986) issued a key reference book that contained many cases vividly illustrating that local users of fisheries, forests and irrigation systems around the world had organized themselves and

overcome the presumed tragedy of the commons. This led to the creation of the CPR database at Indiana University, where attributes of the resource systems and of their users were systematically coded from in-depth case studies that contained sufficient information related to a core set of variables. Thirty inshore fisheries and 47 irrigation systems managed by farmers or by government agencies were carefully coded and statistical analysis undertaken to assess how some of the attributes of resources and of users discussed later in this chapter affect the patterns of organization and performance of these systems (see Schlager 1994; Tang 1994).

We also followed Sen's advice on the importance of game-theoretical analyses as complements to more complex empirical studies (Sen 2002: 207). We first developed an initial game-theoretical analysis of *n* players facing a choice between harvesting from an open-access, common-pool resource (with a quadratic production function as proposed earlier by Gordon 1954) versus allocating effort to an activity that yielded constant returns (e.g. working as a day laborer for a set wage) (Gardner and Ostrom 1991).

Using game theory to formalize the problem enabled us to start an extensive experimental research program by rigorously operationalizing the parameters of the model in an experimental laboratory. When subjects were anonymously placed in a common-pool resource experiment, and not allowed to communicate in any manner, they did substantially overharvest as predicted by Hardin (Walker *et al.* 1990). Simply allowing subjects to engage in face-to-face communication, however, enabled them to increase greatly their individual and joint outcomes, particularly when the opportunity to communicate was repeated over ten or more trials (Ostrom and Walker 1991). To be sure, this form of communication without a third-party enforcer of agreements is considered "cheap talk" in game theory and is posited not to make any difference, let alone the very substantial difference achieved. But allowing subjects to engage in costly punishment and to devise their own rules, monitored by themselves, also greatly increased the joint payoffs, coming close to the optimal level achievable in the formal model (Ostrom *et al.* 1992).

The laboratory experiments did not establish that subjects lacked rationality or were inexperienced (our statistical analysis was based on experiments conducted with subjects who had already experienced our experimental setting), as some critiques of experimental methods have indicated in their effort to retain a narrow model of individual choice in all social settings. Rather, as Sen has argued in his famous article on "Rational Fools" (1977a), the experiments brought to light the fact that the subjects used a different form of reasoning than that presumed in many formal economic theories of human behavior. When subjects had an opportunity for repeated communication, they used this opportunity first to discuss the problem they faced and to ask what joint strategy would get the most money for the group from the experimenters (Ostrom *et al.* 1994: ch. 7). Then they discussed what was the best way of dividing that optimal payoff among members of the group.

Adam Simon and David Schwab (2006) have conducted an in-depth analysis of the verbatim transcripts from these experiments to assess whether participants attempted to increase group identity during the communication rounds. They argue that when individuals see themselves as a member of a potential group, they can use solidarity language to increase the likelihood of other participants seeking group benefits rather than their own short-term benefits. The original transcripts were coded independently by three coders who categorized phrases into multiple categories. In their analysis, Simon and Schwab correlated the number of solidarity words used and the number of defections (individual investment made above the level the group had agreed upon) observed for each period of an experiment. They found that defections were inversely correlated to the number of solidarity words used. Thus, subjects used communication as a way of increasing the sense that they were a group jointly affected by the results they obtained.

III. Toward a Theory of Self-Organization

Evidence from field and experimental research challenges the generalizability of Hardin's "tragedy of the commons" theory. While his theory is successful in predicting outcomes in settings where resource users are alienated from one another or cannot communicate effectively for reasons including the size of the group or their total separation, it does not provide an explanation for settings where resource users are able to engage in collective choice to create their own agreements. Nor does it predict successfully when government ownership will perform well or whether privatization will improve outcomes. After more than three decades of research on whether or not resource users will organize themselves and manage a common-pool resource, it is now possible to provide a theoretical explanation of the factors affecting the likelihood that the users of a common-pool resource will commit themselves to changing the rules from open access to a new set of rules that restricts who can use resource flows, and perhaps also other rules affecting the sustainability of the resource.[4]

Let us assume a set of resource users (U) contemplating a proposed change in the rules related to their use of a common-pool resource (R) (see Ostrom 1990: ch. 6 and Ostrom 1992 for earlier sketches of this theory). Each user i ($i \in U$) has to compare the expected net benefits of harvesting when continuing to use no rules

[4] Earlier, Buchanan and Tullock (1962) developed a theory of constitutional choice regarding the calculus that a group would use to select rules for making collective choices, ranging from an "any one" rule (such as calling the fire department by making one phone call) through to unanimity.

or existing rules that are not working well (*Ro*) to the benefits he or she expects to achieve with a proposed new set of rules (*Rn*). Each user *i* can do this by asking whether his or her evaluation of future benefits (B_i) given the following equation is positive or negative:

$$B_i = Rn_i - Ro_i.$$

If B_i is negative for all users, no one has an incentive to change and one can predict that the resource will remain open-access. If B_i is positive for some users, each of these users needs to estimate three types of costs:

- $C1$: the up-front costs of time and effort spent devising and agreeing upon new rules;
- $C2$: the short-term costs of adopting new harvesting strategies; and
- $C3$: the long-term costs of monitoring and maintaining a self-governed system.

If the sum of these expected costs for all users exceeds the incentive to change, no user will invest the time and resources needed to create new institutions. Thus, if

$$B_i < (C1_i + C2_i + C3_i)$$

for all $i \in U$, no change will occur.

In field settings, not everyone is likely to expect the same costs and benefits from a proposed change. Some may perceive positive benefits after all costs have been taken into account, while others perceive net losses. Consequently, the collective-choice rules used to change the day-to-day operational rules in a group of resource users affect whether an institutional change favored by some and opposed by others will be adopted. Given Arrow's theorem, Sen argues (1999: 354–5), no guarantee exists that any such decisions taken in the field will be Pareto-optimal and meet the full set of conditions that Arrow posited for a good collective-choice rule. As Sen points out, one must recognize that not all collective decisions made in the field are democratic or, even if they are democratic, meet all of Arrow's conditions.

Aoki (2001, 2007a) cogently describes the social games that are likely to exist in any community and how social games and economic games may be linked. In many field studies, resource users draw on the accepted rules that have evolved over time in either social games or political games related to the villages where resource users live. It may be that these rules are used as collective-choice rules to decide on future operational rules related to a common-pool resource.

For any collective-choice arrangement, such as unanimity, majority or a small ruling elite, there is a minimum coalition of users, $K \subset U$, that must agree prior to the adoption of new rules. If for every individual who is a member of U,

$$B_k \leq (C1_k + C2_k + C3_k),$$

no new rules will be adopted. And if at least one coalition $K \subset U$ is such that

$$B_k > (C1_k + C2_k + C3_k)$$

for all members of K, it is feasible for a new set of rules to be adopted. If no minimum winning coalition (given the collective-choice rule in use) evaluates B as greater than C, no new operational rules will be adopted.

If a local chief or other notable has dictatorial powers at the collective-choice level, then only this one person has to estimate that the costs of changing a rule are less than the benefits of a new rule. In this case, of course, there may not be widespread benefits for other members of the group. If the group relies on a larger collective-choice rule and if there are several such coalitions, the question of which coalition will form, and thus which rules will result, is a further theoretical issue that I do not address in this chapter (see Shepsle 1989 and others on coalition-building in collective-choice settings). This analysis is applicable to a situation where a group starts with an open-access set of rules and contemplates adopting its first set of rules limiting access. It is also relevant to the continuing consideration of changing operational rules over time.

The collective-choice rule used to change operational rules in field settings varies from reliance on the decisions made by one or a few leaders, to a formal reliance on majority or super-majority vote, to reliance on consensus or near-unanimity. If there are substantial differences in the perceived benefits and costs of users, it is possible that K will impose a new set of rules on the other users that strongly favors those in the winning coalition and imposes losses or lower benefits on those in the losing coalition (Thompson *et al.* 1988). If the expected benefits from a change in operational rules are not greater than the expected costs for many users, however, the costs of enforcing a change in institutions will be much higher than when most participants expect to benefit from a change in rules over time. Where the enforcement costs are fully borne by the members of K, operational rules that benefit the other users lower the long-term costs of monitoring and sanctioning for a governing coalition. Where external authorities enforce the rules agreed upon by K, the distribution of costs and benefits is more likely to benefit K and may impose costs on the other users.

IV. Variables Affecting the Calculation of Benefits and Costs of Changing Operational Rules

Scholars familiar with the results of field research tend to agree on a set of variables that enhance the likelihood of resource users organizing themselves to avoid the social losses associated with open-access, common-pool resources (Ostrom 1990, 1992; McKean 1992, 2000; Schlager and Ostrom 1992; Tang 1992; Ostrom *et al.* 1994;

Wade 1994; Baland and Platteau 1996; Schweik *et al.* 1997). There is a considerable consensus, drawing on Ostrom (1992: 298–9) and Baland and Platteau (1996: 286–9), that the following attributes of resources and of users are conducive to an increased likelihood that self-governing associations will form.

Attributes of the resource

R1. *Feasible improvement.* Resource conditions have not deteriorated to the point that it is useless to organize, or so underutilized that little advantage results from organizing.

R2. *Indicators.* Reliable and valid indicators of the condition of the resource system are frequently available at a relatively low cost.

R3. *Predictability.* The flow of resource units is relatively predictable.

R4. *Spatial extent.* The resource system is sufficiently small, given the transportation and communication technology in use, that appropriators can develop accurate knowledge of external boundaries and internal microenvironments.

Attributes of the resource users

U1. *Salience.* Appropriators are dependent on the resource system for a major portion of their livelihood.

U2. *Common understanding.* Appropriators have a shared image of how the resource system operates (attributes R1–R4 above) and how their actions affect each other and the resource system.

U3. *Low discount rate.* Appropriators use a sufficiently low discount rate in relation to future benefits to be achieved from the resource.

U4. *Trust and reciprocity.* Appropriators trust each other to keep promises and relate to each other with reciprocity.

U5. *Autonomy.* Appropriators are able to determine access and harvesting rules without external authorities countermanding them.

U6. *Prior organizational experience and local leadership.* Appropriators have learned at least minimal skills of organization and leadership through participation in other local associations or learning about ways that neighboring groups have organized.

V. Integrating Findings from Field Studies with Theory

The attributes of a resource listed above affect both the benefits and costs of institutional change. If resource units are relatively abundant (R1), there are few reasons for users to invest costly time and effort in organizing. If the resource

is already substantially destroyed, the high costs of organizing may not generate substantial benefits. Thus, self-organization is likely to occur only after users observe substantial scarcity (Wade 1994). The danger here, however, is that exogenous shocks leading to a change in the relative abundance of the resource units may occur rapidly and users may not adapt quickly enough to the new circumstances (Libecap and Wiggins 1985).

The presence of frequently available, reliable indicators about the conditions of a resource (R2) affects the capacity of users to adapt relatively quickly to changes that can adversely affect their long-term benefit stream (Moxnes 1998). A resource flow that is highly predictable (R3) is much easier to understand and manage than one that is erratic. This is true for both the users themselves and for public officials who may have acquired management responsibilities for a resource of a particular type in a region (Brock and Carpenter 2007). If the resource flow is unpredictable, it is always difficult for users (or for scientists and government officials) to judge whether changes in the resource stock or flow are due to overharvesting or to random exogenous variables. Unpredictability of resource units in micro-settings, such as private pastures, may lead users to create a larger common-property unit to increase the predictability of resource availability somewhere in the larger unit (Netting 1972; Wilson and Thompson 1993). The spatial extent of a resource (R4) affects the costs of defining reasonable boundaries and then of monitoring them over time.

The attributes of the users themselves also affect their expected benefits and costs. If users do not obtain a major part of their income from a resource (U1), the high costs of organizing and maintaining a self-governing system may not be worth their effort. If users do not share a common understanding of how complex resource systems operate (U2), they will find it extremely difficult to agree on future joint strategies. As Libecap and Wiggins (1985) argue, asymmetric private information about heterogeneous assets may adversely affect the willingness of participants to agree to a reduction in their use patterns before considerable damage is done to a resource. Given the complexity of many common-pool resources—especially multi-species or multi-product resources—understanding how these systems work may be challenging even for those who make daily contacts with the resource. In resources that are highly variable (U3), it may be particularly difficult to understand and to sort out those outcomes stemming from exogenous factors and those resulting from the actions of users. And as Brander and Taylor (1998) have argued, when the resource base itself grows very slowly, population growth may exceed the carrying capacity before participants have achieved a common understanding of the problem they face. Of course, this is also a problem facing officials as well as users. Users with many other viable and attractive options, who thus discount the importance of future income from a particular resource (U3), may prefer to "mine" one resource without spending resources to regulate it (Berkes *et al.* 2006). They simply move on to other resources, becoming "roving bandits" once one is destroyed, assuming there will always be other resources available to them.

Users who trust one another to keep agreements and use reciprocity in their relationships with one another (U4) face lower expected costs in monitoring and sanctioning one another over time. Users who lack trust at the beginning of a process of organizing may be able to build this form of social capital (Coleman 1988) by initially adopting small changes before trying to make major institutional changes. Autonomy (U5) tends to lower the costs of organizing. A group that has little autonomy may find that those who disagree with locally developed rules seek contacts with higher-level officials to undo the efforts of users to achieve their own new rules.

With the legal autonomy to make their own rules, users face substantially lower costs in defending their own rules against other authorities. Prior experience with other forms of local organization (U6) greatly enhances the repertoire of rules and strategies known by local participants as potentially useful for achieving various forms of regulation. Further, users are more likely to agree upon rules whose operation they understand from prior experience than rules introduced by external actors and new to their experience. Given the complexity of many field settings, users face a difficult task in evaluating how diverse variables affect expected benefits and costs over a long period of time. In many cases, it is just as difficult, if not more so, for scientists to make a valid and reliable estimate of total benefits and costs and their distribution (Wilson *et al.* 2007).

The theory presented above and the empirical evidence do *not* imply that most users of common-pool resources will undertake self-governed regulation (as recognized in Berkes 2007 and Meinzen-Dick 2007). There are many settings where the theoretical expectation should be the opposite: the resource will be overused unless efforts are made to change one or more of the variables affecting perceived costs or benefits. Given the number of such variables, many points of external intervention can enhance or reduce the probability of users agreeing upon and following rules that generate higher social returns (Nagendra 2007). Both social scientists and policy-makers have a lot to learn about how these variables operate interactively in field settings (Ostrom 2007).

Many aspects of the macro-institutional structure surrounding a particular setting affect the perceived costs and benefits. Thus, external authorities can do a lot to enhance or reduce the likelihood and performance of self-governing institutions. Further, when the activities of one set of users, U_j, have "spillover effects" on others beyond U_j, external authorities can either facilitate processes that allow multiple groups to solve conflicts arising from negative spillovers or take a more active role in governing particular resources themselves.

Researchers and public officials need to recognize the multiple manifestations of these theoretical variables in the field. Users may be highly dependent on a resource (U1), for example, because they are in a remote location and there are few roads which enable them to leave. Alternatively, they may be located in a central location, but other opportunities are not open to them due to lack of training

or a discriminatory labor market. Users' discount rates in relation to a particular resource (U3) may be low because they have lived for a long time in a particular location and expect that they and their grandchildren will remain in that location, or because they possess a secure and well-defined bundle of property rights to this resource (see Schlager and Ostrom 1992). Reliable indicators of the condition of a resource (R2) may result from activities of the users themselves—such as regularly shearing the wool from sheep (see Gilles and Jamtgaard 1981)—or because of efforts to gather reliable information by users or by external authorities (Blomquist 1992; Blomquist and Ostrom 2008). Predictability of resource units (R3) may result from a clear regularity in the natural environment or because storage evens out the flow of resource units over both good and bad years (Schlager *et al.* 1994). Users may have autonomy to make their own rules (U5) because a national government is weak and unable to exert authority over resources that it formally owns, or because national law formally legitimates self-governance.

When the benefits of organizing are commonly understood by participants to be very great, users lacking many of the attributes conducive to the development of self-governing institutions may nonetheless be able to overcome their liabilities and develop effective agreements. The crucial factor is *not* whether all of the attributes are favorable but the relative size of the expected costs and benefits they generate as perceived by the participants who comprise a winning coalition given the collective-choice rules in use. All of these variables affect the expected costs and benefits of users. It is difficult, however, particularly for outsiders, to estimate their impact on expected costs and benefits, given the difficulty of measuring these variables precisely and weighing them on a cumulative scale.

VI. Puzzles

Despite the consensus concerning the variables most likely to enhance self-organization, many unresolved issues remain about the self-governance of common-pool resources. Two major empirical and theoretical questions concern the effect of size and heterogeneity.

VI.1 Size

The effect of the number of participants on the likelihood of creating and sustaining a self-governing enterprise is unclear. Many theorists argue that the size of a group is negatively related to solving collective-action problems in general because of

its impact on the costs of agreement in the first place (C1) and of monitoring (C3) in the second. Many results from game-theoretical analysis of repeated games conclude that cooperative strategies are more likely to emerge and be sustained in smaller rather than larger groups (see the overview of this literature in Baland and Platteau 1996). Scholars who have studied many self-organized irrigation and forestry institutions in the field have concluded that success will more likely happen in smaller groups (see e.g. Barker *et al.* 1984; Cernea 1989). Other empirical studies have generated mixed results. While most of the 37 farmer-governed irrigation systems studied by Tang (1992: 68) were relatively small, ranging in size from 7 to 300 appropriators, he did not find any statistical relationship within that size range between the number of appropriators or the amount of land being irrigated and performance variables. In Lam's multiple regression analysis of the performance of a much larger set of irrigation systems in Nepal ranging in size up to 475 irrigators, he did not find any significant relationship between either the number of appropriators or the amount of land included in the service area and any of the three performance variables he studied (Lam 1998: 115).

One of the problems with a focus on group size as a key determining factor is that many other variables change as group size increases. If the costs of providing a public good related to the use of a common-pool resource, say a sanctioning system, remain relatively constant as group size increases, then increasing the number of participants brings additional resources that can be drawn upon to provide the benefit enjoyed by all (see Isaac *et al.* 1994). On the other hand, if one is analyzing the transaction costs of arriving at acceptable allocation formulas, group size may well exacerbate the problems of self-governing systems. Given trade-offs among various impacts of size on other variables, a better working hypothesis is that group size has a curvilinear relationship to performance (Agrawal 2000).

VI.2 Heterogeneity

One reason some scholars conclude that only very small groups can organize themselves effectively is because they presume that size is related to the homogeneity of a group. Heterogeneity is also a highly contested variable. For one thing, groups can differ along a diversity of dimensions including their cultural backgrounds, interests and endowments (see Baland and Platteau 1996). Each may operate differently.

If groups from diverse cultural backgrounds share access to a common resource, the key question affecting the likelihood of self-organized solutions is whether the views of the multiple groups concerning the structure of the resource, authority, the interpretation of rules, trust, and reciprocity differ or are similar. In other words, do they share a common understanding of their situation (U2)? New settlers to a region may simply learn and accept the rules of the established group, and their cultural

differences on other fronts do not affect their participation in governing a resource. On the other hand, new settlers are frequently highly disruptive to the sustenance of a self-governing enterprise when they generate higher levels of conflict over the interpretation and application of rules and increase enforcement costs substantially.

When the interests of users differ, achieving a self-governing solution to common-pool resource problems is particularly challenging. Users who possess more substantial economic and political assets may have interests similar to those with fewer assets, or they may differ substantially on multiple attributes. When the more powerful have similar interests, they may greatly enhance the probability of successful organization if they invest their resources in organizing a group and devising rules to govern that group. Those with substantial economic and political assets are more likely to be members of *K* and thus have a bigger impact on decisions about institutional changes.

Differences in the endowments of appropriators can be associated with both extreme levels of conflict and very smooth and low-cost transitions to a sustainable, self-governed system. Heterogeneity of wealth or power may or may not be associated with differences in interests. When those who have more assets share similar interests with those who have fewer, groups may be privileged by having the more powerful take on the higher initial costs of organizing while crafting rules that benefit a large proportion of the appropriators. Users may design institutions that cope effectively with heterogeneities. Thus, when they adopt rules that allocate benefits using the same formulae used to allocate duties and responsibilities, users whose assets differ significantly will tend to agree to and follow such rules (see Design Principles in Ostrom 1990: 90).

Even in a group that differs on many variables, if at least a minimal winning subset of *K* harvesting an endangered but valuable resource is dependent on it (U1), shares a common understanding of its situation (U2), has a low discount rate (U3), trusts one another (U4), and has autonomy to make its own rules (U5), it is more likely that it will estimate the expected benefits of governing the resource to be greater than the expected costs. Whether the rules agreed upon distribute benefits and costs fairly depends both on the collective-choice rule used and the type of heterogeneity existing in the community. Neither size nor heterogeneity are variables with a uniform effect on the likelihood of organizing and sustaining self-governing enterprises. Instead of focusing on size or the various kinds of heterogeneity by themselves, it is important to ask how these variables affect other variables as they affect the cost–benefit calculus of those involved in negotiating and sustaining agreements.[5] Their impact on costs of producing and distributing information is particularly important.

[5] When analyzing the effect of multiple variables, simple mathematical models become very difficult to use, but agent-based models become a powerful tool (see Janssen and Ostrom 2006a, b, c).

VII. Conclusion

Engaging with impossibility results is much to be preferred to oversimplified rejection or despair. We have all learned much from Amartya Sen as he has engaged with a number of impossibility issues for all of us. It is not possible to repay Sen for all he has taught us, but I hope that sharing the experience of engaging with impossibilities is an appropriate form of homage to one of the great scholars of this age.

References

Agrawal, A. (2000), "Small is Beautiful, but is Larger Better? Forest-Management Institutions in the Kumaon Himalaya, India", in C. Gibson, M. McKean and E. Ostrom (eds), *People and Forests: Communities, Institutions, and Governance* (Cambridge, Mass.: MIT Press), 57–86.

Aoki, M. (2001), *Toward a Comparative Institutional Analysis* (Cambridge, Mass.: MIT Press).

—— (2007a), "Linking Economic and Social-Exchange Games: From the Community Norm to CSR", keynote speech delivered at the Conference on Social Capital, Corporate Social Responsibility, and Sustainable Economic Development, University of Turin, July.

—— (2007b), "Three-Level Approach to the Rules of Societal Games: Generic, Substantive, and Operational", Working Paper (Stanford, Calif.: Stanford University).

Arnold, J. E. M., and Stewart, W. C. (1991), *Common Property Resource Management in India*, Tropical Forestry Papers 24 (Oxford: Oxford Forestry Institute).

Arrow, K. (1951), *Social Choice and Individual Values* (New York: Wiley).

—— (1963), *Social Choice and Individual Values*, 2nd edn (New York: Wiley).

Baland, J.-M., and Platteau, J.-P. (1996), *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* (Oxford: Clarendon Press).

Barker, R., Coward, E. W., Jr., Levine, G., and Small, L. E. (1984), *Irrigation Development in Asia: Past Trends and Future Directions* (Ithaca, NY: Cornell University Press).

Barrett, G. W., and Mabry, K. E. (2002), "Twentieth-Century Classic Books and Benchmark Publications in Biology", *Bioscience*, 52: 282–6.

Berkes, F. (1985), "The Common Property Resource Problem and the Creation of Limited Property Rights", *Human Ecology*, 13: 187–208.

—— (2007), "Community-Based Conservation in a Globalized World", *PNAS*, 104(39): 15188–93.

—— Hughes, T. P., Steneck, R. S., Wilson, J. A., and Bellwood, D. R. *et al.* (2006), "Globalization, Roving Bandits, and Marine Resources", *Science*, 311(5767): 1557–8.

Blomquist, W. (1989), *The Performance of Groundwater Management. Volume 8: The Mojave River Basin*. Bloomington: Indiana University, Workshop in Political Theory and Policy Analysis.

Blomquist, W. (1992), *Dividing the Waters: Governing Groundwater in Southern California* (Oakland: ICS Press).

—— and Ostrom, E. (2008), "Deliberation, Learning, and Institutional Change: The Evolution of Institutions in Judicial Settings", *Constitutional Political Economy* (online); available at <www.springerlink.com/content/11j222gx422387g6/fulltext.pdf>.

Brander, J. A., and Taylor, M. S. (1998), "The Simple Economics of Easter Island: A Ricardo–Malthus Model of Renewable Resource Use", *American Economic Review*, 88: 119–38.

Brock, W. A., and Carpenter, S. R. (2007), "Panaceas and Diversification of Environmental Policy", *PNAS*, 104(39): 15206–11.

Bromley, D. (1989), *Economic Interests and Institutions: The Conceptual Foundations of Public Policy* (Oxford: Basil Blackwell).

—— (1991), *Environment and Economy: Property Rights and Public Policy* (Oxford: Basil Blackwell).

—— (1992), "The Commons, Property, and Common-Property Regimes", in D. Bromley *et al.* (eds), *Making the Commons Work: Theory, Practice, and Policy* (San Francisco: ICS Press), 3–16.

Buchanan, J. M., and Tullock, G. (1962), *The Calculus of Consent: The Logical Foundations of Constitutional Democracy* (Ann Arbor: University of Michigan Press).

Cernea, M. M. (1989), *User Groups as Producers in Participatory Afforestation Strategies*, World Bank Discussion Papers 70 (Washington, DC: World Bank).

Ciriacy-Wantrup, S. V., and Bishop, R. C. (1975), " 'Common Property' as a Concept in Natural Resource Policy", *Natural Resources Journal*, 15: 713–27.

Coleman, J. S. (1988), "Social Capital in the Creation of Human Capital", *American Journal of Sociology*, 94 (suppl.): S95–S120.

Cordell, J. C., and McKean, M. (1992), "Sea Tenure in Bahia, Brazil", in D. Bromley *et al.* (eds), *Making the Commons Work: Theory, Practice, and Policy* (San Francisco: ICS Press), 183–205.

Feeny, D. (1988), "Agricultural Expansion and Forest Depletion in Thailand", in J. F. Richard and R. P. Tucker (eds), *World Deforestation in the Twentieth Century* (Durham, NC: Duke University Press), 112–43.

Gardner, R., and Ostrom, E. (1991), "Rules and Games", *Public Choice*, 70: 121–49.

Gilles, J. L., and Jamtgaard, K. (1981), "Overgrazing in Pastoral Areas: The Commons Reconsidered", *Sociologia Ruralis*, 21: 129–41.

Gordon, H. S. (1954), "The Economic Theory of a Common Property Resource: The Fishery", *Journal of Political Economy*, 62: 124–42.

Hardin, G. (1968), "The Tragedy of the Commons", *Science*, 162: 1243–8.

—— (1998), "Extension of 'the Tragedy of the Commons' ", *Science*, 280: 682–3.

Higgs, R. (1996), "Legally Induced Technical Regress in the Washington Salmon Fishery", in L. J. Alston, T. Eggertsson and D. C. North (eds), *Empirical Studies in Institutional Change* (New York: Cambridge University Press), 247–79.

Isaac, R. M., Walker, J., and Williams, A. (1994), "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups", *Journal of Public Economics*, 54: 1–36.

Janssen, M., and Ostrom, E. (2006a), "Adoption of a New Regulation for the Governance of Common-Pool Resources by a Heterogeneous Population", in J.-M. Baland, P. Bardhan

and S. Bowles (eds), *Inequality, Cooperation, and Environmental Sustainability* (Princeton, NJ: Princeton University Press), 60–96.

—— —— (2006b), "Empirically Based, Agent-Based Models", *Ecology and Society*, 11(2) (online); available at <www.ecologyandsociety.org/vol11/iss2/art37>.

—— —— (2006c), "Governing Social-Ecological Systems", in L. Tesfatsion and K. L. Judd (eds), *Handbook of Computational Economics: Agent-Based Computational Economics*, vol. 2. (Amsterdam: Elsevier), 1465–509.

Kiser, L., and Ostrom, E. (1982), "The Three Worlds of Action: A Metatheoretical Synthesis of Institutional Approaches", in E. Ostrom (ed.), *Strategies of Political Inquiry* (Beverly Hills, Calif.: Sage), 179–222.

Lam, W. F. (1998), *Governing Irrigation Systems in Nepal: Institutions, Infrastructure, and Collective Action* (Oakland: ICS Press).

Libecap, G., and Wiggins, S. N. (1985), "The Influence of Private Contractual Failure on Regulation: The Case of Oil Field Unitization", *Journal of Political Economy*, 93: 690–714.

McCay, B. J., and Acheson, J. M., (1987), *The Question of the Commons: The Culture and Ecology of Communal Resources* (Tucson: University of Arizona Press).

McKean, M. (1992), "Management of Traditional Common Lands (*Iriaichi*) in Japan", in D. Bromley *et al.* (eds), *Making the Commons Work: Theory, Practice, and Policy* (San Francisco: ICS Press), 63–98.

—— (2000), "Common Property: What is it, What is it Good for, and What Makes it Work?", in C. Gibson, M. McKean, and E. Ostrom (eds), *People and Forests: Communities, Institutions, and Governance* (Cambridge, Mass.: MIT Press), 27–55.

Meinzen-Dick, R. (2007), "Beyond Panaceas in Water Institutions", *PNAS*, 104(39): 15200–5.

Moxnes, E. (1998), "Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics", *Management Science*, 44: 1234–48.

Nagendra, H. (2007), "Drivers of Reforestation in Human-Dominated Forests", *PNAS*, 104(39): 15218–23.

Netting, R. McC. (1972), "Of Men and Meadows: Strategies of Alpine Land Use", *Anthropological Quarterly*, 45: 132–44.

—— (1981), *Balancing on an Alp: Ecological Change and Continuity in a Swiss Mountain Community* (New York: Cambridge University Press).

NRC (National Research Council) (1986), *Proceedings of the Conference on Common Property Resource Management* (Washington, DC: National Academy Press).

Oakerson, R. J. (1992), "Analyzing the Commons: A Framework", in D. Bromley *et al.* (eds), *Making the Commons Work: Theory, Practice, and Policy* (San Francisco: ICS Press), 41–59.

Olson, M. (1965), *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, Mass.: Harvard University Press).

Ostrom, E. (1965), "Public Entrepreneurship: A Case Study in Ground Water Basin Management", unpubl. Ph.D. diss., University of California, Los Angeles.

—— (1990), *Governing the Commons* (New York: Cambridge University Press).

—— (1992), "The Rudiments of a Theory of the Origins, Survival, and Performance of Common-Property Institutions", in D. Bromley *et al.* (eds), *Making the Commons Work: Theory, Practice, and Policy* (San Francisco: ICS Press), 293–318.

—— (2005), *Understanding Institutional Diversity* (Princeton, NJ: Princeton University Press).

Ostrom, E. (2007), "A Diagnostic Approach for Going beyond Panaceas", *PNAS*, 104(39): 15181–97.

—— Gardner, R., and Walker, J. (1994), *Rules, Games, and Common-Pool Resources* (Ann Arbor: University of Michigan Press).

—— and Walker, J. (1991), "Communication in a Commons: Cooperation without External Enforcement", in T. R. Palfrey (ed.), *Laboratory Research in Political Economy* (Ann Arbor: University of Michigan Press), 287–322.

—— —— and Gardner, R. (1992), "Covenants With and Without a Sword: Self-Governance is Possible", *American Political Science Review*, 86: 404–17.

Ostrom, V., and Ostrom, E. (1977), "Public Goods and Public Choices", in E. S. Savas (ed.), *Alternatives for Delivering Public Services: Toward Improved Performance* (Boulder, Colo.: Westview Press), 7–49.

Schlager, E. (1994), "Fishers' Institutional Responses to Common-Pool Resource Dilemmas", in E. Ostrom, R. Gardner and J. Walker (eds), *Rules, Games, and Common-Pool Resources* (Ann Arbor: University of Michigan Press), 247–67.

—— Blomquist, W., and Tang, S. Y. (1994), "Mobile Flows, Storage, and Self-Organized Institutions for Governing Common-Pool Resources", *Land Economics*, 70: 294–317.

—— and Ostrom, E. (1992), "Property-Rights Regimes and Natural Resources: A Conceptual Analysis", *Land Economics*, 68: 249–62.

Schweik, C. M., Adhikari, K., and Pandit, K. N. (1997), "Land-Cover Change and Forest Institutions: A Comparison of Two Sub-Basins in the Southern Siwalik Hills of Nepal", *Mountain Research and Development*, 17: 99–116.

Sen, A. (1977a), "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, 6: 317–44.

—— (1977b), "Social Choice Theory: A Re-Examination", *Econometrica*, 45: 53–89.

—— (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Oxford University Press).

—— (1984), *Resources, Values, and Development* (Oxford: Blackwell).

—— (1986), "Social Choice Theory", in K. J. Arrow and M. Intriligator (eds), *Handbook of Mathematical Economics*, vol. 3 (Amsterdam: North-Holland), 1073–181.

—— (1993), "Internal Consistency of Choice", *Econometrica*, 61: 495–521.

—— (1999), "The Possibility of Social Choice", *American Economic Review*, 89: 349–78.

—— (2002), *Rationality and Freedom* (Cambridge, Mass.: Harvard University Press).

Shepard, G. (1992), *Managing Africa's Tropical Dry Forests* (London: Overseas Development Institute).

Shepsle, K. A. (1989), "Studying Institutions: Some Lessons from the Rational Choice Approach", *Journal of Theoretical Politics*, 1: 131–49.

Simon, A., and Schwab, D. (2006), "Say the Magic Word: Effective Communication in Social Dilemmas", Working Paper (Bloomington: Indiana University, Workshop in Political Theory and Policy Analysis).

Tang, S. Y. (1992), *Institutions and Collective Action: Self-Governance in Irrigation* (San Francisco: ICS Press).

—— (1994), "Institutions and Performance in Irrigation Systems", in E. Ostrom, R. Gardner and J. Walker (eds), *Rules, Games, and Common-Pool Resources* (Ann Arbor: University of Michigan Press), 225–47.

Thompson, L. L., Mannix, E. A., and Bazerman, M. H. (1988), "Group Negotiation: Effects of Decision Rule, Agenda and Aspiration", *Journal of Personality and Social Psychology*, 54: 86–95.

Thomson, J., Feeny, D., and Oakerson, R. J. (1992), "Institutional Dynamics: The Evolution and Dissolution of Common-Property Resource Management", in D. Bromley *et al.* (eds), *Making the Commons Work: Theory, Practice, and Policy* (San Francisco: ICS Press), 129–60.

Wade, R. (1994), *Village Republics: Economic Conditions for Collective Action in South India* (San Francisco: ICS Press).

Walker, J., Gardner, R., and Ostrom, E. (1990), "Rent Dissipation in a Limited-Access Common-Pool Resource: Experimental Evidence", *Journal of Environmental Economics and Management*, 19: 203–11.

Wilson, J., Yan, L., and Wilson, C. (2007), "The Precursors of Governance in the Maine Lobster Fishery", *PNAS*, 104(39): 15212–17.

Wilson, P. N., and Thompson, G. D. (1993), "Common Property and Uncertainty: Compensating Coalitions by Mexico's Pastoral *Ejidatorios*", *Economic Development and Cultural Change*, 41: 299–318.

CHAPTER 29

AGENTS INTO PRINCIPALS

DEMOCRATIZING DEVELOPMENT IN SOUTH ASIA

REHMAN SOBHAN

I. Introduction: Scope of the Chapter

This chapter attempts to integrate two abiding concerns which have characterized Amartya Sen's work—his commitment to ending poverty and hunger and his belief in the need to build an equitable and just society (Sen 1981; Sen and Drèze 1989; Sen 1999). Since much of the empirical content of Sen's work has been inspired by his ongoing love affair with the people of South Asia, the discussion is contextualized within the specific experience of this region.

The argument underlying this chapter is that the persistence of poverty across South Asia is due to the incapacity of its policy-makers to address the structural sources of the problem. Poverty originates in unequal command over both economic and political resources within society and the unjust nature of a social order

I would like to acknowledge the contribution of Professor M. M. Akash to the preparation of this chapter, and the research assistance of Sanjida Shamsher Elora.

which has created and perpetuates these inequities. We may term these inequities "structural injustice" (Sobhan 2006). Such injustice remains pervasive across the countries of South Asia. The poverty reduction strategies of South Asia have largely failed to address these injustices, so that the region is not only home to the largest number of poor people across the world but its society remains deeply divided, mired in deprivation and riven with tension (Sobhan and Akash 2006). This injustice has compromised the working of its political institutions, so that the crisis of poverty is compounded by the malfunctioning of its democracies.

The prevailing institutional approach to poverty reduction has focused on strengthening the agency role of the poor. South Asia is something of a world leader in non-governmental organizations (NGOs) such as the Grameen Bank (which is actually a commercial bank), BRAC, SEWA and many other such globally recognized institutions which have strengthened the agency capacities of the poor. Government interventions have also supported such initiatives directly or indirectly (see Narayan and Glinskaya 2007). This chapter argues that the poverty reduction strategies across South Asia now need to move beyond strengthening the agency role of the poor through designing policies and building institutions which serve to promote the poor from agents into principals in the development process, through broadening opportunities for asset ownership and enhancing their scope for sharing in the value-addition process (Sobhan 2006). I draw upon the experience of five countries of South Asia: Bangladesh, India, Nepal, Pakistan and Sri Lanka. The remainder of this chapter will be structured around three main topics:

1. South Asia's approach to poverty alleviation
2. The South Asian Poverty Reduction Strategy Papers (PRSPs)
3. Democratizing South Asia's development

II. South Asia's Approach to Poverty Alleviation

II.1 Development Outcomes and Shared Structural Features of the South Asian Countries

There have been some indications of poverty reduction across South Asia in the last two decades but progress is modest and levels of poverty in all countries except Sri Lanka remain unacceptably high (SAARC 2004). The social indicators for the region, as reflected in the trends in the human development indicators (HDIs) as prepared by the United Nations Development Programme (UNDP) in their Human

Development Reports (HDRs), are also poor but indicate quite significant reductions in almost all of the countries (Sri Lanka is an exception but its HDI rank was already quite high compared to its neighbors) (HDR 2005). Some explanation for this apparent paradox is provided by the qualitative content of the social outcomes and indicators of each country as well as their socio-economic policy regimes. In this context it will be instructive to draw upon HDR South Asia 2003, which notes the following commonalities in the development experiences of the South Asian countries:

- Mixed GDP growth rates across the region, with some acceleration in the last ten years except in Nepal (see Sobhan and Akash 2006).
- Increasing inequalities of income in all the economies of the region.
- A deceleration in employment generation.
- Modest reduction in poverty, as defined by the headcount ratio of poverty.
- A relative decline in manufacturing growth rates, especially in the small-scale sector.
- Quality, as well as terms and conditions, of employment has, in general, deteriorated in all South Asian countries.
- All these countries have converged more or less on a common set of policies within the framework of the so-called Structural Adjustment Reforms (SAR) sponsored by the Bretton Woods institutions (BWIs).

These commonalities in South Asia's development experience originated in certain shared structural features of their economy, polity and society, which may be summarized as follows:

- Highly unequal distribution of physical, financial and human assets.
- Political instability originating from various forms of ethnic or communal conflicts as well as regional and interpersonal inequalities in income.
- A high degree of dualism between the organized and unorganized sector.
- The continuing importance of the agricultural sector as a major source of employment.
- The early emergence of a large service sector as one of the largest employers, often acting as the last refuge of the unemployed poor.
- A major segment of the assetless poor engaged in self-employment based on low-productivity micro-activities.

The above evidence on trends in growth and poverty across South Asia indicates that growth remains a necessary but hardly sufficient element in the reduction of poverty. We therefore need to know much more about the sources and quality of this growth to track the causality which links growth to poverty reduction. It may therefore be more useful at this stage to diagnose the development strategies of the respective governments of South Asia and the specific emphasis they have given to poverty reduction. It should be kept in mind that the goal of poverty reduction has been a central concern of South Asia's governments since independence. However,

political and policy commitments to remove poverty have not proved particularly fruitful, so poverty has persisted and, in some countries of the region, has been aggravated (Sobhan and Akash 2006). We will initially look at the response to poverty taken by the South Asian governments acting collectively. We will then review the most recent national strategies reflected in the respective PRSPs of Bangladesh, Nepal, Pakistan, Sri Lanka and the Tenth Five-Year Plan of India.

II.2 The SAARC Agenda for Poverty Alleviation

The First SAARC Initiative on Poverty

The first attempt to address the issue of poverty within the framework of a regional governmental body, the South Asian Association for Regional Cooperation (SAARC), was initiated during the SAARC summit in Colombo in 1991. On the basis of the Colombo Summit decision, the Independent South Asian Commission for Poverty Alleviation (ISACPA) was formed in November 1992, with the following terms of reference (SAARC 1993):

- To diagnose what had gone wrong with past attempts at poverty alleviation.
- To learn positive lessons from the grassroots, where the poor have been mobilized to contribute to economic growth and human development.
- To identify the critical elements in a coherent overall strategy of poverty alleviation in South Asia.

The main findings of the first report of ISACPA, produced in 1993, identified quite a few structural constraints inhibiting the eradication of poverty in this region. The report started with the recognition that South Asia is caught in the trap of low economic growth and poverty perpetuation. It then identified certain forces that perpetuate poverty:

- Population growth.
- In all South Asian countries there are remote undeveloped areas where cost-effective poverty alleviation programs are difficult to implement.
- The growing foreign debt burden for the whole region.
- The emerging global economic environment, as the globalization process is opening up opportunities only for the few while posing a threat for many.
- The reluctance of the developed countries to open up their markets. (ISACPA 1993)

In light of the above diagnosis of poverty the Commission suggested that agricultural development and food security must be maintained because of their relatively greater direct positive impact on the lives of the low-income groups. The report also pointed out that the government's involvement in the development process was excessive and should be replaced by the participation of the poor in the development

process. This recognition of the importance of strengthening the agency role of the poor reflected the strong presence on the Commission of members from organizations engaged in working with the poor at the grassroots. The report recommended that the poor should be mobilized and empowered socially, economically, culturally and politically. A strategy of forming community-based organizations (CBOs) and mobilizing the poor was also launched in various SAARC countries on a pilot basis. But the program, inspired by the ISACPA report, did not originate from any national government or from the SAARC secretariat but was sponsored and coordinated by the UNDP through its regional office in Kathmandu through a cross-country pilot project known as the South Asian Programme for the Alleviation of Poverty (SAPAP). Such an externally sponsored effort at reducing poverty proved unsustainable in most of the SAARC countries and was eventually discontinued by the UNDP (Sobhan 2002). Some features of SAPAP, however, were incorporated in programs sponsored by the government in Andhra state in India (Aiyar *et al.* 2007) and in Pakistan through the National Rural Support Project (Rasmussen *et al.* 2007).

The ISACPA report was adopted at the SAARC summit in Dhaka in 1993, which made a commitment to eliminate poverty across South Asia by 2002. However, following the Dhaka summit, no government made a concerted effort to give substance to the ISACPA report, let alone honor their summit commitment to end poverty (Sobhan and Akash 2006).

During this period, whilst poverty persisted across South Asia, its prevalence across much of the developing world was also emerging as a global concern. More than a decade of structural adjustment reforms (SAR), largely sponsored by the Bretton Woods Institutions, the World Bank and the International Monetary Fund (IMF), had done little to sustain growth or reduce poverty except in the regions of East Asia (Lipumbo 2001). The new development discourse, inspired by the policy failures and the limited impact of development aid over the past two decades, inspired the Millennium Summit of 2000 at the United Nations. The Summit moved on to proclaim the Millennium Development Goals (MDG), which seek to halve extreme poverty across the world by 2015 and also to attain a number of other related social goals (HDR 2005).

Poverty Alleviation Revisited

In view of the changed global perspective on poverty and in view of South Asia's own melancholy progress in reducing poverty, the South Asian heads of states, meeting at the Kathmandu SAARC Summit of 2002, reconstituted the ISACPA and entrusted a new commission with the task of preparing a report outlining a fresh poverty reduction strategy (PRS) for South Asia more attuned to the external realities of the new millennium.

Naturally the following question arose: what specific changes have occurred in South Asia and the world during the 1990s that have made a new strategy necessary? Some of the major changes during the decade of the 1990s, along with a set of new recommendations, were summarized in the 2003 ISACPA report (SAARC 2004). These changes include:

- Trade liberalization: South Asia's average tariff level declined from 65% in the late 1980s to about 30% at the end of the 1990s, and had been further reduced in the last seven years. However, in spite of the extensive reduction in the tariff rate the share of exports from South Asia in global exports had barely increased, from 0.6% in the late 1980s to around 1% at the end of 2005. Most of this incremental rise in exports was contributed by labor-intensive manufactures, which at least brought about some positive diversification in the traditional structure of South Asian exports to other countries.
- Direct foreign investment (DFI) in South Asia had risen significantly during the last two decades. However, its distribution was highly skewed in favor of India, with relatively little going to the smaller countries of the region.
- Remittances from migrant workers had increased exponentially. This had enhanced foreign exchange earnings and personal income in many countries of South Asia.
- New technological breakthroughs, especially in the ICT sector. Again, these gains were mostly concentrated in India, which had emerged as a globally competitive exporter of IT services.

According to the 2003 ISACPA report these positive opportunities for an outward-looking South Asia created prospects for exploiting the ongoing globalization process for the development of the whole region. At the same time the BWIs and other development institutions had also begun to change their approach towards poverty. The international agencies had gradually begun to move towards giving priority to poverty reduction rather than growth as the focus of their development assistance. The BWIs now made it mandatory for all recipients of development assistance to prepare Poverty Reduction Strategy Papers (PRSPs), which had both to reaffirm the country's prioritization of poverty reduction in their development agenda and to demonstrate strong domestic ownership over their planning process.

The 2003 ISACPA report ultimately suggested three strategic priorities:

1. Mobilizing the power of the poor
2. Prudent macroeconomics
3. Mainstreaming the informal sector

The various policies and programs within these three strategic areas involved the following partial departures from the strategy proposed in the 1993 ISACPA report:

- The new approach recognized that the fruits of globalization are unequally shared between the rich and the poor, but saw some prospects of the poor capturing a greater share of the benefits given an increase in the efficiency and productivity of the backward regions and the countries as a whole.
- The new approach especially emphasized rural development so that the benefits of globalization would not remain concentrated in the urban areas.
- The new report emphasized regional cooperation in the field of trade and investment, and urged the South Asian countries to work together to reduce tariffs for better market access to the developed countries, along with seeking increased aid and further debt relief.
- The new report called for prudent macroeconomic policies to ensure higher growth as well as stability. This actually involved a set of policies more in line with the earlier generation of structural adjustment reforms of the Bretton Woods institutions, put in place across South Asia in the 1980s and '90s.
- The report proposes reforms to improve the investment climate.
- The report recognizes that liberalization and deregulation may have various negative consequences for the poor but, according to the new strategy, liberalization is to be accepted. What is required is not trying to stop or reverse liberalization but establishing some safety nets and social protection measures to compensate the short-term losers.
- The report calls for greater investment to ensure more effective education and improvements in socio-economic indicators. Such measures for social improvement and the accompanying safety net measures would be sufficient to ensure higher participation of the poor in the growth process.
- Particular policies were also put forward for supporting informal-sector workers through development of their skills and literacy, ensuring a minimum level of wages and health and safety measures, along with provision for maternity benefits, child care, insurance against old age, illness, etc.
- Various macro-policies were suggested, such as preferential tax and tariffs to support the informal sector, technology and market support, and new laws to facilitate entry into the formal sector.
- To empower people at the grassroots level, the report proposes measures for improving local democracy by delegating real power to local bodies, holding regular free and fair elections at the local level, and eliminating the under-representation of the underclasses in these local bodies.

How far the recommendations of the latest ISACPA report will be operationalized remains to be seen. Three of the five governments which approved the report—Bangladesh, Nepal and Pakistan—are either no longer in office or are mired in political turmoil. In all such moments of political transition SAARC, and particularly its secretariat, tend to lose momentum and less pressing initiatives are put on hold.

III. PRSPs in South Asia

III.1 PRSP Exercises in South Asia

In any case, SAARC as a regional organization can do little to tackle poverty, which remains an exclusively national project. The 2003 ISACPA report was thus intended as a challenge to the respective SAARC governments to demonstrate their commitment to reducing poverty. It also offered suggestions as to how they might share some of their experiences. Both of the ISACPA reports drew upon the national agendas for poverty alleviation in the respective member countries of SAARC. The 2003 ISACPA report was obviously influenced by the PRSPs which were being prepared in various member countries. In the more recent five-year plans of some of these countries, explicit goals and programs for reducing poverty have been incorporated.

The BWI-initiated PRSP process was, at best, a continuum of various ongoing programs of poverty alleviation inherited from the successive five-year plans of the SAARC countries (Helleiner 2000). The point of departure of the PRSP process in South Asia was therefore its explicit focus on poverty reduction and donor influence in the design of the PRSPs. Five out of the seven countries of South Asia have already published their PRSPs: Bangladesh, Bhutan, Nepal, Pakistan and Sri Lanka. India declined to participate in the PRSP exercise, arguing that its Tenth Five-Year Plan (TFYP) document (2002–7) provided a sufficiently comprehensive national strategy for poverty reduction.

III.2 A Comparative Analysis of the South Asian PRSPs

The PRSP Process

If we look carefully into the procedures for formulating PRSPs in four of the South Asian countries we can identify some common features (Sobhan and Akash 2006):

- In all four of these cases the national government undertook the PRSP exercise at the behest of the donor agencies. This experience was common to the PRSP process in most countries in other regions (Lipumbo 2001; Blett and Plant 2002).
- In all these cases the process was initiated from the top under the general guidance of the finance ministry, where bureaucrats and technical experts played a very significant role in preparing the PRSP.
- The donor's inputs were exceedingly crucial in finalizing the PRSP. In all cases the PRSP had been submitted to the donors for their critical approval before it

was given its final shape. This too was common to the PRSP process across the world (Ames and Klugman 2002).

- Contrary to the original mandate laid out by the BWIs (IMF 2001; World Bank 2000) the degree of political commitment and domestic ownership of the PRSP document naturally remains quite low, due to its external inspiration. The PRSP process also suffered from inadequate nationwide participation by the grassroots. To be sure, some civil society organizations (CSOs), community-based organizations (CBOs) and groups of poor people were selected, but numbers were limited. Nor were the participants given any opportunity to share in the decision-making process for finalizing the PRSP. As a result, it was far from clear how much of the grassroots inputs were actually incorporated in the PRSP. The final document thus remained beyond the control of the grassroots contributors and its content largely depended on the preferences and interests of those who selected these participants for the dialog.
- Nowhere in South Asia was there a transparent and serious discussion in the parliament about the premises and the outcome of the PRSP exercise. Nor was it approved (as was the tradition with five-year plans) in the elected parliaments of any country of South Asia.
- In Nepal, Pakistan and Sri Lanka there was a more or less explicit attempt to integrate the PRSP process with the old or ongoing practice of five-year development plans. In Bangladesh the PRSP replaced the five-year plan.
- In all these PRSP exercises hypothetical three-year rolling budgets were prepared, though the future financing of these budgets remains uncertain and tentative.

A Shared Approach to Macroeconomic Reforms

Prior to their PRSP exercise, almost all of the South Asian countries had long been following the BWI-sponsored reform agendas known as Structural Adjustment Reforms (SARs). The PRSPs of the four South Asian countries did not actually constitute any decisive break with their ongoing reform processes (Sobhan and Akash 2006). The three PRSP documents of Nepal, Pakistan and Sri Lanka had explicitly recognized their external policy inheritance. Although the Bangladesh document did not categorically recognize this truth it actually incorporates the old, unfinished or remaining reform agendas of the BWIs in the PRSP whilst the government persists with ongoing SARs outside the PRSP. Though India did not undertake a separate PRSP exercise it is also following a SAR-oriented reform agenda. However, successive regimes in India persist in emphasizing the home-grown nature of their reform process and can afford to do so because of the inner strength of their national economy.

By the turn of the millennium most of the economies of South Asia had either initiated or completed the following macroeconomic reforms in their economies:

- Substantial reduction of state control in terms of price controls and regulation of economic activities.
- Reduction of tax rates in general, with a special emphasis on import duties.
- In order to reduce the fiscal deficit the general policy was to cut public expenditure, especially subsidies to farmers and state-owned enterprises.
- Increasing the domestic prices of utilities, oil/gas, and user fees for various social/public services.
- Liberalization of foreign trade, allowing entry of imported goods/services and capital into the domestic market without any restrictions.
- Financial liberalization and the freeing up of interest rates, as well as reduction of subsidized targeted credits.
- Movement toward market-determined exchange rates.
- Liberalization of the capital account, e.g. easing rules for FDI and access to foreign commercial borrowing by domestic firms.

All the above elements of the SAR process were incorporated into the PRSPs of the four countries as the core macroeconomic basis of their PRSP process. Since most of the above reforms are only tangentially related to the reduction of poverty, the SAR ancestry of the PRSPs is all the more evident.

From PRSP to MDG

All four South Asian PRSPs have also more or less aligned their poverty reduction targets with the Millennium Development Goals, which are to be achieved by 2015. These commonly accepted MDGs, which are also applicable to all the South Asian PRSPs, are as follows:

- Eradication of extreme poverty and hunger: reducing by half the population in poverty by 2015.
- Achieving universal primary education.
- Promoting gender equality: eliminating gender disparity in primary and secondary education.
- Reducing child mortality: reducing under-5 mortality by two-thirds.
- Improving maternal health: reducing maternal mortality by three quarters.
- Combating HIV/AIDS, malaria and other diseases.
- Ensuring environmental sustainability: reducing by half the proportion of people without access to safe water.
- Constructing a worldwide partnership of development.

A detailed discussion on the status of MDGs in the different countries of South Asia is available in the 2003 ISACPA report, *Road Map Towards a Poverty Free South Asia*. Its main conclusion is that, except for Sri Lanka, none of the countries of South Asia would be able to achieve the MDGs by 2015 if the current pace of poverty reduction continued and the PRSPs remained confined to "business as usual". It has been

further calculated that given current trends, the MDG target date for South Asia will have to be reset at 2025.

High Growth Rate

All four South Asian PRSPs aspire to accelerate present growth rates. These different growth targets are then used to project corresponding macroeconomic scenarios for the country, depicting the behavior of broad macro-variables such as rates of investment, savings, government expenditure, foreign aid, inflation rates, employment, etc. No attempt has been made in any of the four countries to link these macroeconomic variables causally to poverty reduction or income-distribution outcomes. Nor is there any attempt to spell out the poverty-reducing impact of specific programs and policies incorporated in the PRSPs or the growth process itself.

Human Resource Development

All of the PRSPs aim to invest in enhancing the human resource base of their countries. In Bangladesh the main mechanism for investing in human resource development is to increase budgetary expenditure in the health and education sector. The two issues of the quality of the health service and education and the shortage of funds for these public services are recognized explicitly in the Bangladesh PRSP (Government of Bangladesh 2005). However, no attention is paid to the economic significance and socio-political implications of the widening disparity in quality of education between the deprived majority and a privately educated elite (Sobhan and Akash 2006). Poor management of public education and health is recognized as an important feature of the public institutions in the human development sector. However, the PRSP does not go beyond supporting the involvement of NGOs and CBOs to ensure better accountability of teachers and doctors and thereby better quality of public services. It is also suggested that current public spending in the education sector be increased from the present level of 2.2% of GDP to 4.4%.

In Nepal elaborate plans were presented for both the health and education sector, with a much greater emphasis on increasing adult literacy to 63% with the help of CBOs and NGOs (Government of Nepal 2003). However, Nepal too does not take into account the issue of growing social disparity in their own human development.

In Pakistan the main goal in the education sector is to ensure the implementation of the ongoing Education Sector Reforms Action Plan (2001–5), which is embedded in the national PRSP (Government of Pakistan 2003). This required an increase in public spending on education from 1.8% of GDP in 2001/2 to 2.5% by 2007/8. In the health sector the main aim is to fight certain diseases such as TB, malaria, Hepatitis B, polio, HIV and AIDS. A program to immunize children against six communicable diseases and a new community-based maternal and child health

program is also to be implemented. Here again, the PRSP is largely silent on the chasm which divides a small, privately educated elite and a large majority served by public services or religious institutions, particularly in the area of education. In Pakistan, this gap is perhaps the most extreme in South Asia and is now contributing to the politically explosive alienation of a poorly educated, impoverished underclass.

For Sri Lanka the main goal for human resource development under the PRSP is to reduce regional and class disparity through enhancing the quality of both education and health services uniformly across the country and across classes (Government of Sri Lanka 2002). Sri Lanka's PRSP has also set the relatively more ambitious target of attaining universal secondary school enrollment by 2010. In the health sector Sri Lanka also aims to ensure better services by introducing more modern and superior technology as well as automation into its public health centers.

Safety Net Provisions

All four South Asian PRSPs have their own safety net programs (Sobhan and Akash 2006):

- Bangladesh has a four-fold approach to its safety net program, aimed especially at those sections of the population who face poverty due to shocks emanating from natural disasters, sudden changes in market prices, changes in government policy, retrenchment policy, etc.
- In Nepal the safety net programs consist of a large number of target-group-oriented expenditure/transfers by the government. The delivery mechanism, however, is dependent on NGOs, CBOs and local government.
- In Pakistan a micro-finance bank has been established to serve the very poor. Other social safety nets currently available to the vulnerable groups in Pakistan include a variety of welfare programs targeted to the poor.
- In Sri Lanka the main issue is to reform Samurdi, the largest welfare program in Sri Lanka, to ensure better targeting, fewer leaks and increased efficiency.

Good Governance

All four PRSPs recognize, at least in words, the necessity of good governance as well as increasing agency through the voice and participation of the poor in the overall process of poverty alleviation (Sobhan and Akash 2006).

- In this regard the position of the Bangladeshi PRSP is quite explicit. It suggests the promotion of good governance through improving implementation capacity, promoting local governance, tackling corruption, enhancing access to justice for the poor, and improving the service delivery mechanism.

- The Nepalese PRSP proposes increased decentralization, civil service reforms, improved financial management and accountability, and various anti-corruption measures for ensuring good governance. Nothing much is said about increasing social mobilization or participation, or the voice of the poorer classes of society.
- Improving governance and devolution is one of the four pillars of the Pakistani PRSP. The elements of this pillar are: devolution, fiscal decentralization, access to justice, police reforms, civil service reforms, pay and pension reform, capacity-building, an anti-corruption strategy, procurement reform, freedom of information, fiscal and financial transparency, and the improvement of statistics. As regards the agency role of the poor, the PRSP confines itself to suggesting the enabling of the community participation of the poor through the legal framework in the village and neighborhood councils and Citizens Community Boards (CCBs).

Structural Issues in the PRSPs

None of the five South Asian countries have addressed the structural dimensions of poverty in their PRSPs or in their recent five-year plans (FYPs) (Sobhan and Akash 2006). All the plans acknowledge the inequitable nature of their respective societies but attempt to deal with the problem through emphasizing the need to enhance the agency role of the less privileged sections of society. The issue of broadening asset ownership has been addressed largely through programs for distribution of public lands, where available, to the landless, improved land-titling as well as reforms, and modernization of the land administration system.

In spite of the inequitable distribution of land and the exploitative nature of tenancy arrangements, no country has introduced provisions for land redistribution or tenancy reforms. In Bangladesh, India and Nepal, lands, bodies of water and forests under the control of the state have traditionally been earmarked for distribution to the resourceless, landless and poor fisherfolk or forest dwellers, some of whom have traditionally enjoyed usufructory rights over particular natural resources. However, much of the public land, water or other natural resources have been illegally appropriated by local elites with access to political patronage (Barkat 2001; Saha 2007). Attempts by the state to recapture these public resources have been frustrated by the same elites.

Asset ownership by the poor has been encouraged in the PRSPs through programs which build upon ongoing initiatives to promote micro-credit. Most of these micro-credit initiatives have originated in actions outside the government through organizations such as the Grameen Bank or the plethora of micro-finance institutions (MFIs) around the region (Zaman 2007). The PRSPs/plans have sought to extend the coverage of micro-credit through public interventions promoting the development of wholesale credit institutions such as the PKSF in Bangladesh or

the Pakistan Poverty Alleviation Fund (PPAF). Such programs have been initially funded by the World Bank or the Asian Development Bank (ADB). In India, national and state-level programs have been designated to direct micro-credit to particular segments of the rural poor either through NGOs or CBOs (Aiyar *et al.* 2007). Attempts to expand the scope of micro-credit to lending for bigger investments, including small business enterprises with a capacity to employ labor as well as to expand agricultural credit to small farmers, have also been emphasized in the PRSPs and the FYPs. Other financial products for the poor, such as insurance against personal loss or crop loss, as well as housing credit for slum dwellers and assetless rural citizens, have also been promoted.

All these initiatives to expand the financial capabilities of the poor have served to enhance their asset base marginally and widen their business opportunities through micro-enterprise. But the level and direction of such lending have never been sufficient, or even intended, to confront the inequities which divide South Asian society between a privileged elite and a resource-poor majority. The inherited structural arrangements, which concentrate ownership of land, corporate assets, access to financial assets in the formal sector, command over the market, and control over the value-addition process have barely been recognized and have thus been left unaddressed in the policy agendas of South Asia. In this respect, the South Asian PRSP process is no different from PRSPs across the world, which reflect the astructural approach to poverty driving the aid/lending programs of the international development agencies such as the World Bank and the ADB (Sobhan and Akash 2005).

IV. From Agents to Principals: Democratizing the Development Agenda

IV.1 Promoting Agency

We have noted that the approach to poverty in South Asia has evolved significantly over the years but that it still remains pervasive across the region. A reduction in poverty has been registered in every country, though how much of this is due directly to conscious public design or the impact of policy reforms remains debatable. The role of non-governmental initiatives in both delivering resources to the poor and promoting their agency role has been one of the conspicuous features of the poverty reduction process in South Asia (Narayan and Glinskaya 2007).

The international development agencies have in recent years taken cognizance of the importance of enhancing the agency role of the poor or resourceless and have incorporated this approach into their policy agendas, where they have emphasized the importance of the role of governance in reducing poverty (see e.g. World Bank 2002). It has been argued that making the poor participants in the design of development projects, the oversight of public expenditure, and the implementation and management of these programs will not only ensure more effective implementation but will alleviate the disempowerment which characterizes their lives (World Bank 2003).

The contemporary approach to poverty reduction has thus moved towards focusing on policy reforms that promote agency through the empowerment of the poor. The objective is to raise the individual consciousness of the poor whilst equipping them with credit or education to enhance their capability to participate in the market economy. These programs have been inspired by the ongoing efforts of various NGOs. Micro-credit is a typical example of promoting agency through focusing on the fiduciary responsibility of the individual poor. This approach to enhancing the agency of the poor at the individual level has been contested by various reformers, who argue that the empowerment of the poor demands collective action through institutions such as cooperatives or self-help groups (SHGs). These reformers argue that in an unequal society, empowerment demands ending the isolation of the poor and encouraging them to draw strength from their numbers (Khan 1998; Sobhan 2002). This is the only way to sever the vertical ties of domination and dependency which characterize their social and working lives.

Development agencies initially gave little recognition to the importance of collective action in the promotion of agency. The collective efforts of the cooperative movements of the 1950s and '60s in South Asia had been largely discredited and superseded by the individual-centered programs of the micro-credit era of the 1980s, which was more in keeping with the belief in individual enterprise and market-centered reforms of the 1980s. Unfortunately, the individual-centered approach to agency was not very useful in promoting investments in local infrastructure projects such as roads, irrigation channels, rural market-places, school housing, etc., which are also essential to poverty reduction. Such projects generate public goods for the community, and so demand collective oversight in their design and management. If the poor are to benefit from such projects through the exercise of their agency function, it is best that they act collectively. The World Bank and other development agencies had for some time been critical of the waste and mismanagement of public agencies who were entrusted with implementing such infrastructure projects (World Bank 2004). These development institutions were persuaded that public investments would be better managed by the exercise of agency by the intended beneficiaries (World Bank 2004). The recognition of social capital as a development variable has also encouraged aid donors to suggest that public service delivery could

be improved by enhancing the agency role of the beneficiaries acting as a collective body.

This transition from the individual to the collective exercise of agency, however, is still to be incorporated into the institutional design of poverty programs as part of a more generalized policy transition. This tension originates in the challenge to the existing hierarchies of local and national power inherent in any moves for collective action by the disempowered, at least in South Asia. Hierarchies are built around the distribution of economic and political power. The power relationship between those who control employment opportunities and those who seek work, between those who own land and those who lease land, or those who sell their primary produce and those who add value to it, cannot be democratized just by building community organizations. Ownership and control over land is profoundly unequal in South Asia. For example, over 50% of rural households in Bangladesh are landless (Saha 2007). Access to cultivable land has an impact both on the functioning of the tenancy market and commodity markets and influences access to formal credit as well as public resources.

Access of a poor woman to micro-credit to buy a cow does not address the problem of who controls the milk market or the distribution chain or value addition associated with the processing of milk. Investing milch cow owners with a sense of agency to seek a better price for their milk may help to improve the price they receive from milk traders. But the primary producer still remains at the bottom of the value chain for milk products and enhancement of their agency role within a market-driven system will not significantly enhance returns on their supply of milk. It is only when the woman as an agent can graduate to the position of a principal, as a possible shareholder in a milk-processing enterprise, that the original micro-credit beneficiary will not only be able to enhance their income but also acquire a stake in an important industry and share in its growth. The circumstances of a member of the Amul Dairy Cooperative in Gujarat, the biggest dairy enterprise in India, and a woman in Bangladesh with a micro-credit loan from BRAC which has enabled her to buy a cow, are qualitatively different (Khandker 2007; Kurien 2007).

IV.2 Emphasizing Principalship

The central message of this chapter is to suggest that the transformation of the position of the poor and resourceless in the social hierarchies of South Asia will be substantively realized when a policy agenda is put in place which enables them to graduate from the role of agents into principals. Such a transformation will require a measure of structural change in the relations of production and distribution. Such a transition from agent to principal will contribute to democratizing the patron–client relationship which prevails between the individual poor and the asset-owning elites who control their lives and livelihoods. By changing the incentive regime

for agents, whether as workers in enterprises or sharecroppers/wage laborers being provided with a permanent stake in the land, will enhance productivity at least as much as transforming public into privately owned enterprises (Sobhan 1993).

Such a process of transformation will require an agenda for change which will need to focus on broadening opportunities for asset ownership by the resourceless and enhancing the share of primary producers in the value-addition chain. Such an agenda need not limit itself to broadening ownership of such assets as land and bodies of water. The scope for substantive income gain and accumulation of wealth is provided by the modern sectors of the economy controlled by corporate entities. The poor should be encouraged to graduate from the ghetto of the micro-sector into the modern economy through provision of opportunities to acquire ownership of corporate assets (Sobhan 2006). These opportunities should be extended not just to collectives of the resourceless but also to workers in corporate enterprises. Poverty and injustice in South Asia will not end just by creating more employment opportunities for young women in Bangladesh to spend the rest of their lives earning $30 per month as casual labor in the export-oriented garment industry. These women too deserve to be provided with a direct stake in the growth of Bangladesh's most dynamic industry. Providing incentive structures as well as financial resources for private entrepreneurs should also be extended to transform their workers into their partners so that the promotion of private enterprise can be a fully inclusive process.

The PRSPs of South Asia should accordingly suggest financial sector reforms which can channel credit from the formal sector to underwrite the acquisition of corporate assets by organizations of the resourceless or workers. Such loans would not be unsound investments. It should be kept in mind that the poor of South Asia have persistently established their creditworthiness through the high recovery rates reported by the MFIs of the region. This may be contrasted with the relatively higher default rates by elite borrowers from the formal financial sector (Sobhan 1992). Specialized equity funds or mutual funds, underwritten by resources from the development agencies, could also be created to channel equity loans to these groups of the resourceless. These equity loans could also be used to leverage the savings of organizations such as Grameen Bank, the more mature NGOs, or federations of SHGs such as have emerged through the Velegu project in Andhra Pradesh (Aiyar *et al.* 2007), to subscribe to initial public offerings (IPOs) of blue chip companies in the corporate sector across South Asia (Sobhan 2002). Dedicated mutual funds specially designed to mediate the saving of the resource-poor can be designed to perform such a role (Sobhan 2006).

Any agenda for building a stake for the resource-poor in the corporate and other growth sectors of the economy would need to build the institutional architecture necessary to empower the weak and marginal through collective action. Some of the bigger NGOs, for example, should consider progressing from aid-driven service delivery organization to market-competitive corporate organizations owned

by their client base of the resourceless. The Grameen Bank is a good example of such an organization, which is largely owned by 7 million resource-poor women. The Grameen Bank has now been able to acquire a 38% equity stake in Grameen Telecom, which provides cellular telecom services to over 20 million subscribers and is the largest corporate enterprise in Bangladesh, with an estimated capitalized value of $3.5 billion (reported by the CEO of Grameen Telecom, *Daily Star*, Dhaka, 21 November 2007).

IV.3 Democratizing Development

Expanding the role of the poor from agent to principal obviously needs to extend into the political arena. The poor of South Asia are particularly disempowered within the prevailing democratic process, where electoral contestation has largely been a rich man's game (Sobhan 2000). The resource-poor as well as women, who between them account for more than half the voting population, have rarely been represented in the legislatures of the region, nor have their concerns been adequately registered by successive elected governments. The privilege of voting out incumbent governments, which has been the case in Bangladesh, India, Nepal, Pakistan and Sri Lanka whenever such opportunities have been provided to voters, has done little to change the social alignment of politics to accommodate the concerns of the less privileged, even though these constituencies constitute a plurality of the electorate. In a democratic process, it is the citizens as the electorate who should play the role of the principal. And it is the elected representatives who should be assigned the role of agents entrusted to act in the interests of their principals. In South Asia, this process appears to have been reversed. Legislatures have increasingly been captured by elites who have appropriated the role of principals through their command over power and patronage. It is the electorate who have been reduced to clients with little capacity or encouragement even to exercise their agency role.

The failure of South Asia's democratic institutions to ensure that the poor and underprivileged, including significant voting constituencies such as women and minorities, exercise their role in the political process as principals reflects the asymmetries in economic power and social influence in these societies. These inequities originated in the injustices of the past and are hardly related to issues of market competitiveness or inherent talent. The dominant influence in perpetuating wide disparities originates largely in the accident of birth into the appropriate family, caste or class and the capacity to convert these inherited privileges into political resources within an undemocratic society.

I have argued that a credible and sustainable agenda for ending poverty must seek to challenge effectively the asymmetries in power which lie at the source of poverty and injustice in South Asia by democratizing development as well as the

institutions of governance. Such a process obviously cannot ignore the underlying political realities of South Asian societies. Nor should it be seen as an argument for disregarding the many successes in reducing poverty and promoting agency across the region (Narayan and Glinskaya 2007; Nayaran and Petesch 2007).

The spirit underlying this essay is to encourage those who are attempting to give substance to Amartya Sen's lifelong struggle to end poverty across South Asia by seeking to enhance individual freedoms and demonstrating the social commitment to do so (Sen 1999). I have argued that to enhance freedoms we need to correct the structural injustices which constrain freedom and thereby create poverty. Such an approach will demand a qualitative change in the design of poverty alleviation strategies, which should move beyond the prevailing programs for promoting income gain for the poor and enhancing their agency role. The challenge before us is to redesign the institutional architecture so that it can elevate the deprived and marginalized to the role of principals, where they may aspire to own corporate assets, become major players in national and global markets, and eventually emerge as political masters in the legislatures and institutions of local governance across South Asia. Such concrete manifestations of structural change will truly enhance those freedoms so cherished by Amartya Sen.

References

Aiyar, Ankelswamy, *et al.* (2007), "Empowerment through Self Help Groups", in World Bank (ed.), *Ending Poverty in South Asia* (Washington, DC: World Bank).

Ames, Brian, Bhatt, Gita, and Plant, Mark (2002), "Taking Stock of Poverty Reduction Efforts", *Finance and Development*, 39(2) (June); available at <www.imf.org/external/pubs/ft/fandd/2002/06/ames.htm>.

—— and Klugman, Jeni (2002), *Review of the PRSP Approach—Early Experiences with IPRSPs and Full PRSPs,*

Amjad, Rashid (2007), *The Musharaf Development Strategy: Will It Deliver?* Lahore Journal of Development Studies, Lahore

Centre for Policy Dialogue (2004), *Report of the Regional Seminar on the Follow-Up of SAARC Summit Decisions concerning the Report of the ISACPA* (Dhaka: Centre for Policy Dialogue).

—— (2007), *State of the Bangladesh Economy, 2005–2006* (Dhaka: Centre for Policy Dialogue).

Chen, S., and Ravallion, Martin (2004), *How Have the World's Poorest Fared since the Early 1980's?*, Working Paper (Washington, DC: Development Research Group, World Bank).

Government of Bangladesh (2005), *Unlocking the Potential: National Strategy for Accelerated Poverty Reduction* (Dhaka: Planning Commission).

Government of India (2002), *Tenth Five Year Plan 2002–2007* (Delhi: Government of India).

Government of Nepal (2003), *The Tenth Plan: Poverty Reduction Strategy Paper 2002–2007* (Kathmandu: Planning Commission).

Government of Pakistan (2003), *Accelerating Economic Growth and Reducing Poverty: The Road Ahead* (Government of Pakistan, Islamabad: Ministry of Finance).

Government of Sri Lanka (2002), *Regaining Sri Lanka: Vision and Strategy for Accelerated Development* (Colombo: Government of Sri Lanka).

Helleiner, Gerald K. (2000), "External Conditionality, Local Ownership and Development", in Jim Freedman (ed.), *Transforming Development: Foreign Aid for A Changing World* (Toronto: University of Toronto Press).

IMF (2001), *Strengthening Country Ownership of Fund-Supported Programmes*, (Washington, DC: IMF).

—— (2002), *The Modalities of Conditionality: Further Conditions* (Washington, DC: IMF).

—— and IDA (2003), *Poverty Reduction Strategy Paper: Progress in Implementation* (Washington, DC: IMF).

Khan, A. H. (1998), *Orangi Pilot Project: Reminiscences and Reflections* (Karachi: Oxford University Press).

Khandkar, Nasreen (2007), *Strengthening the Capacity of the Poor to Compete in the Market Place*, SACEPS-CPD Project, Research Monograph 9 (Dhaka: Centre for Policy Dialogue).

Kurien, Verghese (2007), "India's Milk Revolution", in World Bank (ed.), *Ending Poverty in South Asia* (Washington, DC: World Bank).

Lipumbo, I. H. (2001), *Conditionality and Ownership: A View From the Periphery* (Washington, DC: IMF).

Mahbubul Haq Development Centre (2003), *Human Development in South Asia 2003* (Karachi: Oxford University Press).

Narayan, Deepa, and Glinskaya, Elena (eds) (2007), *Ending Poverty in South Asia: Ideas that Work* (Washington, DC: World Bank).

—— and Petesch, P. (2007a), "Agency Opportunity Structure and Poverty Escapes", in Narayan and Petesch (2007b).

—— —— (eds) (2007b), *Moving Out of Poverty: Cross-Disciplinary Perspectives on Mobility*, vol. 1 (Washington, DC: World Bank).

Rasmussen, Stephen F., *et al.* (2007), *Scaling up Rural Support Programmes in Pakistan* in Narayan and Glinskaya (2007).

SAARC (1993), *Report of the Independent South Asian Commission on Poverty Alleviation* (Kathmandu: SAARC Secretariat).

—— (2004), *Report of the Independent South Asian Commission on Poverty Alleviation. Our Future, Our Responsibility: Road Map towards a Poverty-Free South Asia* (Kathmandu: SAARC Secretariat).

Saha, Bimal K. (2007), *Land Reform and Peasant Movements in Bangladesh*, SACEPS/CPD Project, Research Monograph 10 (Dhaka: Centre for Policy Dialogue).

Sen, A. K. (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation* (Oxford: Oxford University Press).

—— (1999), *Development as Freedom* (New York: Knopf).

—— and Drèze, Jean (1989), *Hunger and Public Action* (Oxford: Clarendon Press).

Sobhan, Rehman (ed.) (1992), *Debt Default to the Development Finance Institutions: The Crisis of State-Sponsored Entrepreneurship in Bangladesh* (Dhaka: University Press).

Sobhan, Rehman (1993), *Agrarian Reform and Social Transformation: Preconditions for Development* (London: Zed Press).

—— (2000), "Problems of Governance in South Asia: An Overview", in V. A. Pai Panandiker (ed.), *Problems of Governance in South Asia* (Delhi: Konark Press).

—— (2002), *The South Asian Poverty Alleviation Programme* (Kathmandu: UNDP).

—— (2006), "Poverty as Injustice: Refocusing the Policy Agenda", in James K. Boyce, Stephen Cullenberg, Prasanta K. Patnaik and Robert Pollin (eds), *Human Development in the Era of Globalization: Essays in Honor of Keith G. Griffin* (Cheltenham: Edward Elgar).

—— and Akash, M. M. (2005), *Poverty Reduction Strategies of the International Community: The Scope for Structural Change*, SACEPS/CPD Project, Research Monograph 1 (Dhaka: Centre for Policy Dialogue).

—— —— (2006), *Poverty Reduction Strategies in South Asia*, SACEPS/CPD Project, Research Monograph 2 (Dhaka: Centre for Policy Dialogue).

UNDP (2005), *Human Development Report 2005* (New York: UNDP).

World Bank (1994), *Governance: The World Bank's Experience* (Washington, DC: World Bank).

—— (2003), *World Development Report 2004: Making Services Work for Poor People* (Washington, DC: World Bank).

—— (2005), *World Development Report 2006: Equity and Development* (Washington, DC: World Bank).

Zaman, Hassan (2007), "Microfinance in Bangladesh", in World Bank (ed.), *Ending Poverty in South Asia* (Washington, DC: World Bank).

Index of Names

Index of Subjects